BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

~~IMAGES ARE BEST AVAILABLE COPY.~~

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 03:28:19 ; Search time 222.28 Seconds
(without alignments)
347.126 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 989696

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	14	15.6	20	20	AAZ05048	PCR primer used to
c	2	14	15.6	26	19	AAV07952	Helicobacter pylor
c	3	14	15.6	26	19	AAV07922	Helicobacter pylor
c	4	14	15.6	27	19	AAV07937	Helicobacter pylor
c	5	14	15.6	47	21	AAZ68842	Human map-related
c	6	13	14.4	17	16	AAQ92084	Renilla reniformis
	7	13	14.4	20	20	AAX95307	PCR primer used to
	8	13	14.4	23	22	AAH27610	Human lipoprotein
	9	13	14.4	24	22	AAH55939	Human SCN1A PCR-SS
c	10	13	14.4	26	15	AAQ68537	B.thuringiensis 33
	11	13	14.4	26	21	AAC67312	Alzheimer's diseas
	12	13	14.4	26	21	AAC67342	Alzheimer's diseas
c	13	13	14.4	31	19	AAV07957	Helicobacter pylor
	14	13	14.4	34	16	AAT15377	AZF A5F cosmid clo
c	15	13	14.4	35	22	AAF61207	Coxsackie virus B
	16	13	14.4	36	22	AAF90560	Dissociation trans
c	17	13	14.4	38	17	AAT34130	I3L promoter-HTLV-
c	18	13	14.4	39	13	AAQ35630	HIV-2 env 3' fragm
c	19	13	14.4	39	14	AAQ35353	PCR primer HIV2B2.
	20	13	14.4	39	21	AAA99138	Plasmid pOmpTTc PC
c	21	13	14.4	40	14	AAQ48717	Insecticidal prote
c	22	13	14.4	40	21	AAZ61322	Primer 3A used to
	23	13	14.4	40	21	AAZ61323	Primer 3B used to
c	24	13	14.4	41	19	AAV51009	Maize polymorphic
c	25	13	14.4	41	19	AAV51011	Maize polymorphic
c	26	13	14.4	50	16	AAT25074	Human gene signatu
c	27	13	14.4	50	20	AAX52184	Synthetic plasmid
	28	13	14.4	50	20	AAX52048	Synthetic plasmid
c	29	12	13.3	17	18	AAX70046	Human flt1 VEGF re
c	30	12	13.3	17	18	AAX70047	Human flt1 VEGF re
	31	12	13.3	17	20	AAA21453	Integrin alpha 6 s
	32	12	13.3	17	20	AAA21454	Integrin alpha 6 s
	33	12	13.3	17	20	AAA21455	Integrin alpha 6 s
c	34	12	13.3	17	21	AAF03088	Hammerhead ribozym
c	35	12	13.3	17	21	AAF03089	Hammerhead ribozym
c	36	12	13.3	17	21	AAF03090	Hammerhead ribozym
c	37	12	13.3	17	21	AAF03091	Hammerhead ribozym
	38	12	13.3	18	21	AAA08921	Human survivin DNA
c	39	12	13.3	19	15	AAQ62806	Env gene 5' primer
c	40	12	13.3	19	21	AAZ69625	Human biallelic ma
c	41	12	13.3	19	21	AAZ75126	Human biallelic ma
	42	12	13.3	20	11	AAQ05907	HIV mRNA translati
c	43	12	13.3	20	18	AAT80813	Staphylococcus aur
c	44	12	13.3	20	19	AAV72697	Corn kernel oil co
	45	12	13.3	20	20	AAX92327	PCR primer used to

ALIGNMENTS

RESULT 1
AAZ05048
ID AAZ05048 standard; DNA; 20 BP.
XX
AC AAZ05048;
XX
DT 07-OCT-1999 (first entry)
XX
DE PCR primer used to amplify an ORF of Chlamydia trachomatis.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal uretritis; epidymitis; cervicitis; salpingitis; PCR primer;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS Synthetic.
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX

PT Genome sequence of Chlamydia trachomatis
XX
PS Disclosure; Page 1738; 1755pp; English.
XX
CC PCR primers AAZ01426-Z06209 were used to amplify open reading frames
CC (ORFs) of the genome of Chlamydia trachomatis (see AAZ01425). These ORFs
CC encode polypeptides (see AAY36754-Y37949) which can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the
CC invention may be of use in treating these diseases.
XX
SQ Sequence 20 BP; 5 A; 5 C; 5 G; 5 T; 0 other;

Query Match 15.6%; Score 14; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 tattgttggagcaa 67
||||||||||||||
Db 4 tattgttggagcaa 17

RESULT 2
AAV07952/c
ID AAV07952 standard; DNA; 26 BP.
XX
AC AAV07952;
XX
DT 02-FEB-1999 (first entry)
XX
DE Helicobacter pylori polypeptide GHPO 1414 5' DNA primer.
XX
KW GHPO 1414; infection; gastritis; ulcer; vaccine; diagnosis;
KW therapy; PCR; primer; ss.
XX
OS Synthetic.
OS Helicobacter pylori.
XX
PN WO9843479-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06421.
XX
PR 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX
PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J;
XX
DR WPI; 1998-568251/48.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastroduodenal diseases
XX
PS Claim 5; Page 145; 184pp; English.
XX
CC This 5' primer is used with a 3' primer (see AAV07954) in the PCR
CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic
CC DNA in order to obtain DNA (see AAV07921) encoding the unprocessed
CC form of a 76 kDa polypeptide (see AAW73032) designated GHPO 1414.
CC The isolated polynucleotide, and encoded polypeptide, can be used
CC to develop vaccines for the treatment and prevention of Helicobacter
CC infections.
XX
SQ Sequence 26 BP; 14 A; 5 C; 4 G; 3 T; 0 other;

Query Match 15.6%; Score 14; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 ttttcatgttttct 53
||||||||||||||
Db 18 TTTTCATGTTTTCT 5

RESULT 3
AAV07922/c
ID AAV07922 standard; DNA; 26 BP.
XX
AC AAV07922;
XX
DT 02-FEB-1999 (first entry)
XX
DE Helicobacter pylori polypeptide GHPO 386 5' DNA primer.
XX
KW GHPO 386; infection; gastritis; ulcer; vaccine; diagnosis; therapy;
KW PCR; primer; ss.
XX
OS Synthetic.
OS Helicobacter pylori.
XX
PN WO9843479-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06421.
XX
PR 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX
PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J;
XX
DR WPI; 1998-568251/48.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastroduodenal diseases
XX
PS Claim 5; Page 137; 184pp; English.
XX
CC This 5' primer is used with a 3' primer (see AAV07924) in the PCR
CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic
CC DNA in order to obtain DNA (see AAV72001
CC) encoding the unprocessed
CC form of a 76 kDa polypeptide (see AAW73022) designated GHPO 386. The
CC isolated polynucleotide, and encoded polypeptide, can be used to
CC develop vaccines for the treatment and prevention of Helicobacter
CC infections.
XX
SQ Sequence 26 BP; 15 A; 5 C; 4 G; 2 T; 0 other;

Query Match 15.6%; Score 14; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 ttttcatgttttct 53
||||||||||||||
Db 20 TTTTCATGTTTTCT 7

```
RESULT   4
AAV07937/c
ID   AAV07937 standard; DNA; 27 BP.
XX
AC   AAV07937;
XX
DT   02-FEB-1999  (first entry)
XX
DE   Helicobacter pylori polypeptide GHPO 896 5' DNA primer.
XX
KW   GHPO 896; infection; gastritis; ulcer; vaccine; diagnosis;
KW   therapy; PCR; primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
XX
PN   WO9843479-A1.
XX
PD   08-OCT-1998.
XX
PF   31-MAR-1998;   98WO-US06421.
XX
PR   01-APR-1997;   97US-0834666.
PR   01-APR-1997;   97US-0831310.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX
PI   Al-Garawi A,  Kleanthous H,  Lissolo L,  Miller C,  Tomb J;
XX
DR   WPI; 1998-568251/48.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastroduodenal diseases
XX
PS   Claim 5; Page 141; 184pp; English.
XX
CC   This 5' primer is used with a 3' primer (see AAV07939) in the PCR
CC   amplification of Helicobacter, e.g. Helicobacter pylori, genomic
CC   DNA in order to obtain DNA (see AAV07916) encoding the unprocessed
CC   form of a 76 kDa polypeptide (see AAW73027) designated GHPO 896.
CC   The isolated polynucleotide, and encoded polypeptide, can be used
CC   to develop vaccines for the treatment and prevention of Helicobacter
CC   infections.
XX
SQ   Sequence 27 BP; 14 A; 5 C; 4 G; 4 T; 0 other;


  Query Match             15.6%;  Score 14;  DB 19;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      40 ttttcatgttttct 53
           ||||||||||||||
Db      18 TTTTCATGTTTTCT 5


RESULT   5
AAZ68842/c
ID   AAZ68842 standard; DNA; 47 BP.
XX
AC   AAZ68842;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:3195.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,G)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 3; Page 912; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 13 A; 13 C; 6 G; 15 T; 0 other;


  Query Match             15.6%;  Score 14;  DB 21;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 9.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 aatgaaagcaagtg 88
           ||||||||||||||
Db      43 AATGAAAGCAAGTG 30


RESULT   6
AAQ92084/c
ID   AAQ92084 standard; cDNA; 17 BP.
XX
AC   AAQ92084;
XX
DT   07-JAN-1996  (first entry)
XX
DE   Renilla reniformis luciferase DNA probe-1.
XX
KW   Luciferase; enzyme; bioluminescence; luminescence; label; DNA probe;
KW   antibody; oligonucleotide; ss.
XX
OS   Synthetic.
XX
PN   US5418155-A.
XX
```

PD 23-MAY-1995.
XX
PF 29-DEC-1989; 89US-0458952.
XX
PR 29-DEC-1989; 89US-0458952.
PR 20-AUG-1992; 92US-0933017.
PR 17-JUN-1993; 93US-0079700.
PR 14-DEC-1993; 93US-0167650.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Cormier MJ, Lorenz WW;
XX
DR WPI; 1995-199740/26.
XX
PT New recombinant Renilla luciferase polypeptide - used as a
PT luminescent tag, partic in bio-luminescence assays and for the prodn
PT of antibodies
XX
PS Disclosure; Fig. 4; 18pp; English.
XX
CC This 17-mer oligonucleotide DNA probe, along with Probe-2 (AAQ92085)
CC are used to screen an R. reniformis cDNA library to isolate cDNA
CC encoding Renilla luciferase. The luciferase was then expressed
CC using E. coli.
XX
SQ Sequence 17 BP; 6 A; 0 C; 2 G; 9 T; 0 other;

Query Match 14.4%; Score 13; DB 16; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aaaaaaattattt 16
||||||||||||
Db 15 AAAAAAATTATTT 3

RESULT 7
AAX95307
ID AAX95307 standard; DNA; 20 BP.
XX
AC AAX95307;
XX
DT 13-SEP-1999 (first entry)
XX
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; PCR primer; ss.
XX
OS Synthetic.
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1737; Disclosure; 1912pp; English.
XX
CC AAX91991-X97517 represent PCR primers used to amplify open reading
CC frames and other nucleic acid sequences from the genome of
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
CC disease such as pneumonia and bronchitis and is thought to be a
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC containing C. pneumoniae nucleotides sequences can also be used as
CC immunogenic compositions, especially where the vector directs the
CC expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 20 BP; 5 A; 5 C; 5 G; 5 T; 0 other;

Query Match 14.4%; Score 13; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 tgaaagcaagtgc 89
||||||||||||
Db 8 tgaaagcaagtgc 20

RESULT 8
AAH27610
ID AAH27610 standard; DNA; 23 BP.
XX
AC AAH27610;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human lipoprotein 105 PCR primer 2.
XX
KW Human; lipoprotein 105; cytostatic; anti-HIV; antiinflammatory;
KW immunomodulatory; cytochrome c structural domain; cancer;
KW haemopathy; human immunodeficiency virus; HIV; immunological disease;
KW inflammation; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200140483-A1.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-CN00503.
XX
PR 29-NOV-1999; 99CN-0124136.
XX
PA (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2001-374839/39.
XX
PT Human lipoprotein 105 containing cytochrome c structural domain and
PT encoded polynucleotide, applicable in diagnosis and treatment of
PT malignant tumor, hemopathy, HIV infection, immunological diseases and
PT various inflammation -
XX
PS Example 2; Page 12; 40pp; Chinese.
XX
CC The invention relates to an isolated polypeptide of human lipoprotein
CC 105 containing a cytochrome c structural domain. The polypeptide
CC comprises a 951 amino acid sequence given in the specification, or its
CC fragment, analogue or derivative. The polypeptide and encoded
CC polynucleotide are useful in the diagnosis and treatment of malignant
CC tumours, haemopathy, HIV infection, immunological diseases and various
CC inflammatory diseases. The present sequence is a primer which was
CC used to isolate the polynucleotide encoding the polypeptide of the
CC invention.
XX

```
SQ   Sequence 23 BP; 5 A; 1 C; 2 G; 15 T; 0 other;

  Query Match           14.4%;  Score 13;  DB 22;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     33 ttttatgttttca 45
          |||||||||||||
Db      1 ttttatgttttca 13


RESULT   9
AAH55939
ID   AAH55939 standard; DNA; 24 BP.
XX
AC   AAH55939;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Human SCN1A PCR-SSCP PCR primer SEQ ID NO:183.
XX
KW   Human; epilepsy; chromosome 2; SCN1A; SCN2A; SCN3A; identification;
KW   diagnosis; mutation; chromosome 2q23-q31; neurological disorder;
KW   anticonvulsant; neuroprotective; PCR primer; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200138564-A2.
XX
PD   31-MAY-2001.
XX
PF   24-NOV-2000; 2000WO-CA01404.
XX
PR   26-NOV-1999;   99US-0167623.
XX
PA   (UYMC-) UNIV MCGILL.
XX
PI   Rouleau GA,  LaFreniere RG,  Rochefort D,  Cossette P,  Ragsdale D;
XX
DR   WPI; 2001-355945/37.
XX
PT   Determining a predisposition to epilepsy and/or development of epilepsy
PT   comprises determining the genotype of SCN1A, SCN2A and/or SCN3A, or a
PT   DNA variant, equivalent, or mutation which shows a linkage
PT   disequilibrium -
XX
PS   Example 3; Fig 2; 268pp; English.
XX
CC   The present invention describes a method (M1) of determining an
CC   individual's predisposition to epilepsy and/or development of epilepsy,
CC   as well as predicting the individual's response to medication. The
CC   method comprises determining the genotype of at least one gene selected
CC   from SCN1A, SCN2A or SCN3A, or a DNA variant, equivalent, or mutation
CC   which shows a linkage disequilibrium. SCN1a, SCN2A and SCN3A are all
CC   sodium channel genes located on chromosome 2. The idiopathic generalised
CC   epilepsy (IGE) gene is more specifically localised on chromosome
CC   2q23-q31. Compounds identified as modulators of the biological activity
CC   of SCN1A, SCN2A or SCN3A proteins or genes, are useful for treating
CC   epilepsy or other neurological disorders. They have anticonvulsant and
CC   neuroprotective activities. AAH55763 to AAH56164 and AAB99674 to
CC   AAB99679 represent SCN1A, SCN2A, and SCN3A cDNAs, gene fragments, PCR
CC   primers, oligonucleotides and proteins given in the exemplification of
CC   the present invention.
XX
SQ   Sequence 24 BP; 2 A; 5 C; 3 G; 14 T; 0 other;

  Query Match           14.4%;  Score 13;  DB 22;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     34 tttatgttttcat 46
          |||||||||||||
Db      8 tttatgttttcat 20


RESULT  10
AAQ68537/c
ID   AAQ68537 standard; cDNA; 26 BP.
XX
AC   AAQ68537;
XX
DT   13-FEB-1995  (first entry)
XX
DE   B.thuringiensis 33kD delta-endotoxin N-terminal probe.
XX
KW   Insecticidal protein; delta-endotoxin; crystal; Coleoptera;
KW   Lepidoptera; Bacillus thuringiensis; ss.
XX
OS   Synthetic.
XX
PN   WO9413785-A.
XX
PD   23-JUN-1994.
XX
PF   13-DEC-1993;   93WO-US12144.
XX
PR   15-DEC-1992;   92US-0991073.
XX
PA   (NOVO ) NOVO-NORDISK ENTOTECH INC.
XX
PI   Adams LF,  Liu C,  Lufburrow PA,  Thomas MD;
XX
DR   WPI; 1994-217865/26.
XX
PT   New Bacillus thuringiensis strains - which produce new
PT   delta-endotoxin cpds used for the control of Lepidopteran and
PT   Coleopteran insect pests.
XX
PS   Example 9; Page 26; 47pp; English.
XX
CC   The N-terminal sequence of a 33kD delta-endotoxin isolated
CC   from B.thuringiensis EMCC0075 (NRRL B-21019) or EMCC0076 (NRRL B-
CC   21020) was determined (see AAR59764). Based on this sequence a 26mer
CC   oligonucleotide was designed for use as a probe (AAQ68537) for cloning
CC   the delta-endotoxin gene.
XX
SQ   Sequence 26 BP; 13 A; 2 C; 2 G; 9 T; 0 other;

  Query Match           14.4%;  Score 13;  DB 15;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     30 aatttttatgttt 42
          |||||||||||||
Db     16 AATTTTTATGTTT 4


RESULT  11
AAC67312
ID   AAC67312 standard; DNA; 26 BP.
XX
AC   AAC67312;
XX
DT   14-FEB-2001  (first entry)
XX
DE   Alzheimer's disease-linked mitochondrial SNP PCR primer #12.
XX
KW   Human; mitochondrial genome; single nucleotide polymorphism; SNP;
KW   Alzheimer's disease; mtDNA; PCR primer; ss.
XX
OS   Homo sapiens.
XX
```

PN WO200063441-A2.
XX
PD 26-OCT-2000.
XX
PF 19-APR-2000; 2000WO-US10906.
XX
PR 20-APR-1999; 99US-0130447.
PR 22-OCT-1999; 99US-0160901.
XX
PA (MITO-) MITOKOR.
XX
PI Herrnstadt C, Davis RE;
XX
DR WPI; 2000-672748/65.
XX
PT Diagnosing a subject at the risk for or having Alzheimer's disease
PT comprises determining at least one single nucleotide polymorphism in
PT mitochondrial DNA associated with the disease in the sample from the
PT subject -
XX
PS Example 2; Page 34; 89pp; English.
XX
CC The present invention describes a novel method for determining the risk
CC of or diagnosing Alzheimer's disease using single nucleotide
CC polymorphisms (SNPs) present in an individual's mitochondrial DNA
CC (mtDNA). In additon, the SNPs identified can be used to identify agents
CC suitable for use in treating Alzheimer's disease. Sequences
CC AAC67301-C67610 are PCR primers used to demonstrate the method of the
CC invention.
XX
SQ Sequence 26 BP; 8 A; 2 C; 5 G; 11 T; 0 other;

Query Match 14.4%; Score 13; DB 21; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 atgttttcatgtt 49
||||||||||||
Db 7 atgttttcatgtt 19

RESULT 12
AAC67342
ID AAC67342 standard; DNA; 26 BP.
XX
AC AAC67342;
XX
DT 14-FEB-2001 (first entry)
XX
DE Alzheimer's disease-linked mitochondrial SNP PCR primer #42.
XX
KW Human; mitochondrial genome; single nucleotide polymorphism; SNP;
KW Alzheimer's disease; mtDNA; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200063441-A2.
XX
PD 26-OCT-2000.
XX
PF 19-APR-2000; 2000WO-US10906.
XX
PR 20-APR-1999; 99US-0130447.
PR 22-OCT-1999; 99US-0160901.
XX
PA (MITO-) MITOKOR.
XX
PI Herrnstadt C, Davis RE;
XX
DR WPI; 2000-672748/65.
XX
PT Diagnosing a subject at the risk for or having Alzheimer's disease
PT comprises determining at least one single nucleotide polymorphism in
PT mitochondrial DNA associated with the disease in the sample from the
PT subject -
XX
PS Example 4; Page 38; 89pp; English.
XX
CC The present invention describes a novel method for determining the risk
CC of or diagnosing Alzheimer's disease using single nucleotide
CC polymorphisms (SNPs) present in an individual's mitochondrial DNA
CC (mtDNA). In additon, the SNPs identified can be used to identify agents
CC suitable for use in treating Alzheimer's disease. Sequences
CC AAC67301-C67610 are PCR primers used to demonstrate the method of the
CC invention.
XX
SQ Sequence 26 BP; 8 A; 2 C; 5 G; 11 T; 0 other;

Query Match 14.4%; Score 13; DB 21; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 atgttttcatgtt 49
||||||||||||
Db 7 atgttttcatgtt 19

RESULT 13
AAV07957/c
ID AAV07957 standard; DNA; 31 BP.
XX
AC AAV07957;
XX
DT 02-FEB-1999 (first entry)
XX
DE Helicobacter pylori polypeptide GHPO 386 5' DNA primer.
XX
KW GHPO 386; infection; gastritis; ulcer; vaccine; diagnosis; therapy;
KW PCR; primer; ss.
XX
OS Synthetic.
OS Helicobacter pylori.
XX
PN WO9843479-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06421.
XX
PR 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX
PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J;
XX
DR WPI; 1998-568251/48.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastroduodenal diseases
XX
PS Example 3.B; Page 63; 184pp; English.
XX
CC This 5' primer was used with a 3' primer (see AAV07958) in the PCR
CC amplification of Helicobacter pylori strain ORV2001 genomic DNA in
CC order to obtain DNA (see AAV72001
CC) encoding a 76 kDa polypeptide (see
CC AAW73022) designated GHPO 386. The primer pair includes a 5' clamp
CC and BamHI and XhoI restriction enzyme reocgnition sequences for
CC cloning purposes. The PCR product was ligated into vector pET28a,
CC and recombinant polypeptide was expressed as a histidine-tagged
CC fusion protein in E. coli host cells. The polypeptide can be used

```
CC   to develop vaccines for the treatment and prevention of Helicobacter
CC   infections.
XX
SQ   Sequence 31 BP; 14 A; 4 C; 6 G; 7 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 19;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 tttcatgttttct 53
           |||||||||||||
Db      31 TTTCATGTTTTCT 19


RESULT  14
AAT15377
ID   AAT15377 standard; DNA; 34 BP.
XX
AC   AAT15377;
XX
DT   20-MAR-1996  (first entry)
XX
DE   AZF A5F cosmid clone exon 9/intron 9 boundary.
XX
KW   Azoospermia factor; AZF; male infertility; YRRM gene;
KW   Y-chromosome; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1..17
FT                   /*tag= a
FT                   /number= 9
FT                   /note= "Represents bases 1093-1109 of human YRRM gene"
FT   intron          18..34
FT                   /*tag= b
FT                   /number= 9
FT                   /note= "Constitutes 5' end of approx. 450 bp intron 9"
XX
PN   WO9511300-A2.
XX
PD   27-APR-1995.
XX
PF   24-OCT-1994;   94WO-GB02344.
XX
PR   07-JUL-1994;   94GB-0013760.
PR   22-OCT-1993;   93GB-0021857.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Chandley AC,  Cooke HJ,  Hargreave TB,  Kun M,  Sharkey AM;
XX
DR   WPI; 1995-170221/22.
XX
PT   Nucleic acid encoding the human azoospermia factor, and probes and
PT   antibodies specific for the sequence and encoded polypeptide - may
PT   be used in the clinical diagnosis of male infertility
XX
PS   Disclosure; Fig 3; 40pp; English.
XX
CC   The intron/exon boundaries of a human YRRM gene encoding
CC   azoospermia factor were detd. by comparison of the sequences
CC   found in cosmid A5F DNA obtd. from a Y-chromosome specific cosmid
CC   library and the cDNA clone MK5 (AAQ87655) obtd. from an adult human
CC   testis library.
XX
SQ   Sequence 34 BP; 9 A; 3 C; 7 G; 15 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 16;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      24 tactgtaattttt 36
           |||||||||||||
Db      19 tactgtaattttt 31


RESULT  15
AAF61207/c
ID   AAF61207 standard; DNA; 35 BP.
XX
AC   AAF61207;
XX
DT   25-MAY-2001  (first entry)
XX
DE   Coxsackie virus B serotype B3 associated PCR primer SEQ ID 4.
XX
KW   Gene therapy; infectious virion; cardioactive; cardiac muscle; trophic;
KW   cardiac disease; cardiac myocyte; PCR primer; ss.
XX
OS   Coxsackievirus.
XX
PN   DE19939095-A1.
XX
PD   22-FEB-2001.
XX
PF   18-AUG-1999;   99DE-1039095.
XX
PR   18-AUG-1999;   99DE-1039095.
XX
PA   (UYTU-) UNIV TUEBINGEN EBERHARD-KARLS.
XX
PI   Kuepper J,  Meyer R,  Meyer-Ficca M,  Kandolf R;
XX
DR   WPI; 2001-227548/24.
XX
PT   Recombinant RNA comprising heterologous gene in Coxsackie viral genome,
PT   useful in gene therapy, specifically for targeting of cardiac myocytes
PT   -
XX
PS   Example 3; Column 10; 12pp; German.
XX
CC   This invention describes a novel recombinant RNA molecule (I), at least
CC   partly translatable in a target cell which comprises: (a) the
CC   non-infectious genome (A) of Group B Coxsackie virus (CVB), particularly
CC   serotype B3; and (b) at least one foreign gene (II) that can be developed
CC   for a selected function in the target cell, e.g. for gene therapy. The
CC   invention also describes (1) a recombinant infectious virions (V),
CC   derived from (A) and containing (I); (2) a plasmid vector containing the
CC   DNA sequence (III) for (I), under control of a promoter; (3) a helper
CC   construct for complementing the coding sequence exchanged by (I); (4)
CC   producing (V); (5) producing the plasmids of (2); (6) producing helper
CC   constructs of (3); (7) a kit containing the vector of (2) or the helper
CC   construct of (3); (8) a DNA molecule (IV) containing at least one coding
CC   sequence for (I); (9) a kit containing (IV); (10) a kit for performing
CC   methods (5) or (6); (11) a DNA construct that encodes (I) and can persist
CC   (and is transcribed) in target cells, but is preferably not replicable;
CC   (12) a recombinant virus (RV), particularly adeno or retro, that encodes
CC   (I) and is expressed after infection into a target cell to produce a
CC   cytoplasmic replicon that is continuously replenished; (13) producing
CC   recombinant DNA viruses or virions having a DNA genome that lacks a
CC   specific gene function, in which this function is provided from a
CC   recombinant vector system with a RNA genome. The products of the
CC   invention have cardioactive activity and can be used for gene therapy.
CC   (I) Is used to produce gene therapy vectors, particularly plasmids or
CC   virions, and these vectors are used for specific transfer to cardiac
CC   muscle, for diagnosis, prevention or treatment of cardiac disease, either
CC   congenital or acquired. (I) Are also used to complement vectors that lack
CC   particular gene sequences, particularly vectors derived from DNA viruses.
CC   Vectors based on (I) transfer genes to cardiac myocytes without
CC   immunological or other side effects. The RNA genome can replicate,
CC   providing efficient gene transfer and long-term expression of the
CC   therapeutic gene. CBV is naturally trophic for heart muscle and since it
CC   does not produce DNA during its life cycle, overcomes the danger that
```

```
CC   foreign genes will become integrated in the target cell genome. By using
CC   (II) to replace part of the viral coding region, large (II) sequences may
CC   be accommodated. (I) is easily packaged in CVB capsid proteins.
XX
SQ   Sequence 35 BP; 14 A; 2 C; 7 G; 12 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 22;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 aaattatttcaaa 20
          |||||||||||||
Db     27 AAATTATTTCAAA 15


RESULT  16
AAF90560
ID   AAF90560 standard; DNA; 36 BP.
XX
AC   AAF90560;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Dissociation transposon 3' flanking nested primer 3A.
XX
KW   ET1158 gene; GT6839 gene; ET5262 gene; herbicide; screening;
KW   herbicide tolerance; transgenic plant; crop protection;
KW   Dissociation; transposon; maize; PCR primer; ss.
XX
OS   Zea mays.
XX
PN   WO200144277-A2.
XX
PD   21-JUN-2001.
XX
PF   14-DEC-2000; 2000WO-EP12748.
XX
PR   16-DEC-1999;   99US-0465040.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Wegrich Glover L,  Budziszewski GJ,  Levin JZ,  Zhou Q;
XX
DR   WPI; 2001-398122/42.
XX
PT   New herbicide target genes encoding proteins having ET1158, GT6839 or
PT   ET5262 activity, for identifying an inhibitor of protein activity -
XX
PS   Example 1; Page 39; 67pp; English.
XX
CC   The present sequence is that of primer 3A, which is 1 of a set of
CC   3 nested primers (see also AAF90561 and AAF90562) homologous to
CC   the 3' DS3 region of Dissociation transposon. Arbitrary degenerate
CC   primers (see AAF90551-56) were used to prime Arabidopsis thaliana
CC   genomic DNA flanking the site of a Ds transposon insertion. The
CC   degenerate primers were used in combination with 2 sets of 3,
CC   nested, transposon-specific primers homologous to the DS3 region or
CC   DS5 region (see AAF90557-59) of the Ds element, which lie at the
CC   outermost ends of the transposon. Low- and high-stringency PCR
CC   amplifications were performed using the TAIL-PCR protocol. DNA
CC   fragments were produced which corresponded to the genomic DNA that
CC   was directly adjacent to the transposon insertion. Sequence
CC   analysis of PCR products from tagged seedling lethal lines ET1158,
CC   GT6839 and ET5262 identified 3 novel genes (see AAF90548-50) each
CC   of which was essential for Arabidopsis seedling growth and
CC   development. The essentiality of the genes provides a means of
CC   discovering new herbicides. Screening assays for identifying
CC   inhibitors that are potential herbicides are provided. The
CC   invention is also applied to the development of herbicide tolerant
CC   plants, and plant tissues, seeds and cells.
XX
SQ   Sequence 36 BP; 6 A; 4 C; 11 G; 15 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 22;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     36 tatgttttcatgt 48
          |||||||||||||
Db     21 tatgttttcatgt 33


RESULT  17
AAT34130/c
ID   AAT34130 standard; cDNA; 38 BP.
XX
AC   AAT34130;
XX
DT   21-OCT-1996  (first entry)
XX
DE   I3L promoter-HTLV-I fusion primer MW116.
XX
KW   Avipox virus; poxvirus; vector; vaccine; attenuation; HTLV-I
KW   human lymphotropic virus; safety; ALVAC; NYVAC; primer; PCR;
KW   polymerase chain reaction; vaccinia virus; retrovirus; ss.
XX
OS   Synthetic.
XX
PN   WO9621727-A1.
XX
PD   18-JUL-1996.
XX
PF   16-JAN-1996;   96WO-US00547.
XX
PR   13-JAN-1995;   95US-0372664.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Franchini G,  Gallo RC,  Paoletti E,  Tartaglia J;
XX
DR   WPI; 1996-342282/34.
XX
PT   Attenuated recombinant pox-viruses expressing HTLV antigens - for
PT   safe vaccination against HTLV infection, and for therapy and
PT   diagnosis
XX
PS   Example 12; Page 96; 165pp; English.
XX
CC   Primers MW093 (AAT34127) and MW110 (AAT34128) were used to generate
CC   a 100 bp PCR fragment, PCR-HTLV18, contg. the I3L promoter fused to
CC   the 5'-end of the HTLV-I envelope gene, using pMM102 as template.
CC   Primers MW113 (AAT34129) and MW116 (AAT34130) were used to generate a
CC   1,500 bp PCR fragment, PCR-HTLV21, contg. the 3' end of the I3L
CC   promoter fused to the entire HTLV-I envelope gene, using p17-SST as
CC   template.  MW093 and MW116 were then used as to amplify  PCR-HTLV18
CC   and PCR-HTLV21.  The I3L-promoted envelope gene was cloned between
CC   canarypox virus C5 flanking arms and also between vaccinia HA
CC   flanking arms, and a NYVAC recombinant expressing HTLV-I env was
CC   generated.  This effectively primed protective immune responses
CC   against retrovirus challenge.
XX
SQ   Sequence 38 BP; 15 A; 6 C; 8 G; 9 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 17;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     26 ctgtaatttttat 38
          |||||||||||||
Db     25 CTGTAATTTTTAT 13


RESULT  18
```

```
AAQ35630/c
ID   AAQ35630 standard; DNA; 39 BP.
XX
AC   AAQ35630;
XX
DT   24-FEB-1993  (first entry)
XX
DE   HIV-2 env 3' fragment primer HIV2B2.
XX
KW   NYVAC; recombinant; HIV-2; Copenhagen vaccine; vaccinia virus;
KW   virulence factors; deletion loci; recipient loci; env; amplify; PCR;
KW   polymerase chain reaction; gp120; transfection; Vero cells; envelope;
KW   cell surface; ss.
XX
OS   Synthetic.
XX
PN   WO9215672-A.
XX
PD   17-SEP-1992.
XX
PF   09-MAR-1992;   92WO-US01906.
XX
PR   07-MAR-1991;   91US-0666056.
PR   11-JUN-1991;   91US-0713967.
PR   06-MAR-1992;   92US-0847951.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Cox WI,  De Taisne C,  Francis J,  Gettig RR,  Johnson GP;
PI   Limbach KJ,  Norton EK,  Paoletti E,  Perkus ME,  Pincus SE;
PI   Riviere M,  Tartaglia J,  Taylor J;
XX
DR   WPI; 1992-331718/40.
XX
PT   Vaccine comprises recombinant, attenuated pox-virus - use for
PT   vaccinating against viral infections such as rabies, hepatitis B,
PT   HIV, HSV, EBV, CMV, mumps etc.
XX
PS   Disclosure; Page 156; 455pp; English.
XX
CC   The sequences given in AAQ35624-32 were used in the construction of
CC   NYVAC recombinants expressing HIV-2 gene products.  NYVAC is a
CC   Copenhagen vaccine strain of vaccinia virus which has been modified by
CC   deletion of six non-essential regions of the genome encoding known or
CC   potential virulence factors.  The deletion loci were engineered as
CC   recipient loci for the insertion of foreign genes.  The HIV-2 env
CC   sequence was isolated in two portions by polymerase chain reaction
CC   (PCR).  These fragments were then ligated also by PCR.  The HIV-2
CC   gp120 gene was also isolated by PCR.  These HIV-2 genes were
CC   transfected into NYVAC which could then be cultured in Vero cells.
CC   The envelope proteins were found to be present on the cell surface of
CC   cells transformed with the recombinant NYVAC.  See AAQ35501-864.
XX
SQ   Sequence 39 BP; 15 A; 6 C; 7 G; 11 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 13;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT  19
AAQ35353/c
ID   AAQ35353 standard; DNA; 39 BP.
XX
AC   AAQ35353;
XX
DT   18-MAY-1993  (first entry)
XX
DE   PCR primer HIV2B2.
XX
KW   Human immunodeficiency virus; amplification; env; ISSY strain; ss.
XX
OS   Synthetic.
XX
PN   WO9222641-A.
XX
PD   23-DEC-1992.
XX
PF   12-JUN-1992;   92WO-US05107.
XX
PR   14-JUN-1991;   91US-0715921.
PR   11-JUN-1992;   92US-0897382.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Cox WI,  Paoletti E,  Tartaglia J;
XX
DR   WPI; 1993-018128/02.
XX
PT   Modified recombinant virus with inactivated non-essential genetic
PT   functions - comprises e.g. vaccinia or avipox virus, used as HIV
PT   vaccine
XX
PS   Example 4; Page 52; 159pp; English.
XX
CC   The 3' portion of the HIV-2 env gene was derived by PCR.  In this
CC   reaction a 270 bp fragment was amplified using oligonucleotides
CC   HIV2B1 and HIV2B2 using pISSY-KPN as template.  The prod. was
CC   digested with BamHI and XbaI to yield a 150 bp fragment which was
CC   engineered to contain a T5NT sequence known to be recognised as
CC   vaccinia virus early transcription termination signal, following
CC   the termination codon TAA.
CC   See also AAQ35328-437.
XX
SQ   Sequence 39 BP; 15 A; 6 C; 7 G; 11 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 14;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT  20
AAA99138
ID   AAA99138 standard; DNA; 39 BP.
XX
AC   AAA99138;
XX
DT   19-JAN-2001  (first entry)
XX
DE   Plasmid pOmpTTc PCR primer SEQ ID NO:15.
XX
KW   OmpT protease; cleavage; fusion protein; membrane protease;
KW   natriuretic; Escherichia coli; PCR primer; ss.
XX
OS   Escherichia coli.
OS   Synthetic.
XX
PN   WO200052193-A1.
XX
PD   08-SEP-2000.
XX
PF   03-MAR-2000; 2000WO-JP01309.
XX
PR   04-MAR-1999;   99JP-0057731.
XX
PA   (SUNR ) SUNTORY LTD.
```

XX
PI Okuno K, Yabuta M, Ohsuye K;
XX
DR WPI; 2000-579291/54.
XX
PT Controlled cleavage of peptides by OmpT protease by amino acid
PT substitution for ensuring cleavage only at desired site in fission of
PT fusion proteins -
XX
PS Example 2; Fig 9; 144pp; Japanese.
XX
CC The present invention describes a method for regulating the cleavage
CC sites of polypeptides by OmpT protease by preventing cleavage at
CC unwanted sites by converting the amino acid residue at position +1 to
CC the site to a specifically defined amino acid (where the residue at
CC position -1 to the site is Lys or Arg), and/or converting the residue
CC at position -4 and/or -6 to a specifically defined amino acid. Also
CC described is a method for the fission of a fusion protein to give a
CC desired polypeptide by cleavage with OmpT protease, where the fusion
CC protein has a linker peptide inserted between the desired polypeptide
CC and the other part of the fusion protein. The fusion protein may be
CC prepared by expression of DNA encoding in a suitable host cell such as
CC Escherichia coli. OmpT protease is a membrane protease of Escherichia
CC coli which cleaves peptide chains at a two-residue sites in which both
CC residues are basic amino acids such as arginine or lysine. The methods
CC can be used for the efficient preparation of undegraded desired
CC polypeptides such as natriuretic peptide by OmpT protease cleavage after
CC recombinant expression as a fusion protein. AAA99127 to AAA99177 and
CC AAB23946 to AAB24018 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 39 BP; 8 A; 9 C; 4 G; 18 T; 0 other;

Query Match 14.4%; Score 13; DB 21; Length 39;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 ctgtaatttttat 38
||||||||||||||
Db 15 ctgtaatttttat 27

RESULT 21
AAQ48717/c
ID AAQ48717 standard; DNA; 40 BP.
XX
AC AAQ48717;
XX
DT 25-MAR-1994 (first entry)
XX
DE Insecticidal protein gene 3' primer.
XX
KW Caulobacter; plasmid; insecticidal protein; Bacillus thuringiensis;
KW Bacillus sphaericus; larva; mosquitoe; Culex; Anopheles; Psorophoa;
KW Mansonia; Aedes; ss.
XX
OS Synthetic.
XX
PN JP05211866-A.
XX
PD 24-AUG-1993.
XX
PF 05-JUN-1991; 91JP-0160963.
XX
PR 06-JUN-1990; 90JP-0148444.
XX
PA (SILM-) SILMARAN SO TANABAL.
XX
DR WPI; 1993-298916/38.
XX
PT Expression of insecticidal protein - by transforming Caulobacter
PT with plasmid contg. gene coding for insecticidal protein
XX
PS Disclosure; Page 24-25; 27pp; Japanese.
XX
CC Two primers (AAQ48716-17) are described for the isolation of
CC Bacillus thuringiensis israelensis DNA. The sequence of one
CC 27-mer primer (AAQ48716), however is missing from the specification.
CC Caulobacter transformed with a plasmid contg. a gene encoding
CC insecticidal protein derived from Bacillus thuringiensis or
CC Bacillus sphaericus will proliferate in aq. environment.
CC They may be consumed by larvae of mosquitoes and are lethal to
CC Culex, Anopheles, Psorophoa, Mansonia and Aedes.
XX
SQ Sequence 40 BP; 19 A; 6 C; 6 G; 9 T; 0 other;

Query Match 14.4%; Score 13; DB 14; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 tttttatgttttc 44
|||||||||||||
Db 15 TTTTTATGTTTTC 3

RESULT 22
AAZ61322/c
ID AAZ61322 standard; DNA; 40 BP.
XX
AC AAZ61322;
XX
DT 19-JUN-2000 (first entry)
XX
DE Primer 3A used to construct a synthetic bar gene.
XX
KW bar gene; phosphinothricin acetyltransferase; phosphinothricin toxicity;
KW protein expression; vaccine; haemoglobin; enzyme; primer; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
XX
PN WO200007431-A1.
XX
PD 17-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17806.
XX
PR 03-AUG-1998; 98US-0095163.
PR 03-AUG-1998; 98US-0095167.
PR 15-DEC-1998; 98US-0112257.
PR 29-APR-1999; 99US-0131611.
PR 11-JUN-1999; 99US-0138764.
XX
PA (RUTF) UNIV RUTGERS STATE NEW JERSEY.
XX
PI Maliga P, Kuroda H, Khan MS;
XX
DR WPI; 2000-205525/18.
XX
PT New recombinant DNA constructs, for expressing high levels of
PT heterologous protein in plastids of higher plants, includes promoter, a
PT leader sequence and a downstream box element -
XX
PS Example 7; Page 69; 164pp; English.
XX
CC Primers AAZ61218-45 were used to construct a synthetic Streptomyces
CC hygroscopicus bar gene. The bar gene encodes phosphinothricin
CC acetyltransferase, which provides protection from phosphinothricin
CC toxicity. The synthetic gene has improved containment and enhanced
CC expression in plant plastids, and is used to produce recombinant DNA
CC constructs of the invention. The specification describes recombinant DNA
CC constructs for expressing heterologous proteins (e.g. bar gene product)
CC in the plastids of higher plants. The DNA constructs comprise a 5'
CC regulatory region which includes a promoter element, a leader sequence

```
CC   and a downstream box element operably linked to a coding region of the
CC   heterologous protein. The chimeric regulatory region enhances
CC   translational efficiency of an mRNA molecule encoded by the DNA
CC   construct. The DNA constructs are used for producing transformed
CC   monocot and dicot plants having high levels of heterologous protein
CC   expression. They can be used to drive expression of proteins with
CC   agronomic, industrial or pharmaceutical importance, including production
CC   of vaccines, healthcare products like human haemoglobin, industrial or
CC   household enzymes. Plants which can be transformed with the constructs
CC   of the invention include maize, millet, sorghum, sugar cane, rice,
CC   wheat, barley, oat, rye or turf grass.
XX
SQ   Sequence 40 BP; 19 A; 6 C; 5 G; 10 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 21;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      18 aaagtttactgta 30
           |||||||||||||
Db      33 AAAGTTTACTGTA 21


RESULT  23
AAZ61323
ID   AAZ61323 standard; DNA; 40 BP.
XX
AC   AAZ61323;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Primer 3B used to construct a synthetic bar gene.
XX
KW   bar gene; phosphinothricin acetyltransferase; phosphinothricin toxicity;
KW   protein expression; vaccine; haemoglobin; enzyme; primer; ss.
XX
OS   Synthetic.
OS   Streptomyces hygroscopicus.
XX
PN   WO200007431-A1.
XX
PD   17-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17806.
XX
PR   03-AUG-1998;   98US-0095163.
PR   03-AUG-1998;   98US-0095167.
PR   15-DEC-1998;   98US-0112257.
PR   29-APR-1999;   99US-0131611.
PR   11-JUN-1999;   99US-0138764.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Maliga P,  Kuroda H,  Khan MS;
XX
DR   WPI; 2000-205525/18.
XX
PT   New recombinant DNA constructs, for expressing high levels of
PT   heterologous protein in plastids of higher plants, includes promoter, a
PT   leader sequence and a downstream box element -
XX
PS   Example 7; Page 69; 164pp; English.
XX
CC   Primers AAZ61218-45 were used to construct a synthetic Streptomyces
CC   hygroscopicus bar gene. The bar gene encodes phosphinothricin
CC   acetyltransferase, which provides protection from phosphinothricin
CC   toxicity. The synthetic gene has improved containment and enhanced
CC   expression in plant plastids, and is used to produce recombinant DNA
CC   constructs of the invention. The specification describes recombinant DNA
CC   constructs for expressing heterologous proteins (e.g. bar gene product)
CC   in the plastids of higher plants. The DNA constructs comprise a 5'
CC   regulatory region which includes a promoter element, a leader sequence
CC   and a downstream box element operably linked to a coding region of the
CC   heterologous protein. The chimeric regulatory region enhances
CC   translational efficiency of an mRNA molecule encoded by the DNA
CC   construct. The DNA constructs are used for producing transformed
CC   monocot and dicot plants having high levels of heterologous protein
CC   expression. They can be used to drive expression of proteins with
CC   agronomic, industrial or pharmaceutical importance, including production
CC   of vaccines, healthcare products like human haemoglobin, industrial or
CC   household enzymes. Plants which can be transformed with the constructs
CC   of the invention include maize, millet, sorghum, sugar cane, rice,
CC   wheat, barley, oat, rye or turf grass.
XX
SQ   Sequence 40 BP; 8 A; 5 C; 9 G; 18 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 21;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      18 aaagtttactgta 30
           |||||||||||||
Db      28 aaagtttactgta 40


RESULT  24
AAV51009/c
ID   AAV51009 standard; DNA; 41 BP.
XX
AC   AAV51009;
XX
DT   04-JAN-1999  (first entry)
XX
DE   Maize polymorphic marker S41G2/G5-2 DNA.
XX
KW   Polymorphic marker; allele-specific; primer; probe; amplification;
KW   hybridisation; plant; hybrid certification; genetic contribution;
KW   progeny; back-cross; hybrid; ancestry; maize; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   variation       21
FT                   /*tag= a
FT                   /replace= "g"
FT                   /note= "polymorphism"
XX
PN   WO9824796-A1.
XX
PD   11-JUN-1998.
XX
PF   01-DEC-1997;   97WO-US21782.
XX
PR   07-MAR-1997;   97US-0813507.
PR   02-DEC-1996;   96US-0032069.
XX
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Landry BS,  Lemieux B,  Murigneux A,  Sapolsky RJ;
XX
DR   WPI; 1998-333252/29.
XX
PT   Brassica species allele-specific oligonucleotide probes and primers
PT   - useful for plant breeding
XX
PS   Claim 1; Page 43; 65pp; English.
XX
CC   This DNA sequence is a region of a Zea mays genome which contains a
CC   polymorphic marker. This sequence can be used in the construction of
CC   allele-specific primers and probes for amplification or hybridisation,
CC   e.g. to determine common or disparate ancestry between 2 or more plants,
CC   to monitor the genetic contribution of an ancestral plant, to trace the
CC   progeny of proprietary plants, in certification of a hybrid plant or to
CC   identify the progeny of a back-crossed plant with an ancestral plant.
```

```
XX
SQ   Sequence 41 BP; 10 A; 9 C; 6 G; 16 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 19;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 aaaaaattatttc 17
          |||||||||||||
Db     36 AAAAAATTATTTC 24


RESULT  25
AAV51011/c
ID   AAV51011 standard; DNA; 41 BP.
XX
AC   AAV51011;
XX
DT   04-JAN-1999  (first entry)
XX
DE   Maize polymorphic marker S41G2/G4-1 DNA.
XX
KW   Polymorphic marker; allele-specific; primer; probe; amplification;
KW   hybridisation; plant; hybrid certification; genetic contribution;
KW   progeny; back-cross; hybrid; ancestry; maize; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   variation       21
FT                   /*tag=  a
FT                   /replace= "g"
FT                   /note= "polymorphism"
XX
PN   WO9824796-A1.
XX
PD   11-JUN-1998.
XX
PF   01-DEC-1997;   97WO-US21782.
XX
PR   07-MAR-1997;   97US-0813507.
PR   02-DEC-1996;   96US-0032069.
XX
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Landry BS,  Lemieux B,  Murigneux A,  Sapolsky RJ;
XX
DR   WPI; 1998-333252/29.
XX
PT   Brassica species allele-specific oligonucleotide probes and primers
PT   - useful for plant breeding
XX
PS   Claim 1; Page 43; 65pp; English.
XX
CC   This DNA sequence is a region of a Zea mays genome which contains a
CC   polymorphic marker. This sequence can be used in the construction of
CC   allele-specific primers and probes for amplification or hybridisation,
CC   e.g. to determine common or disparate ancestry between 2 or more plants,
CC   to monitor the genetic contribution of an ancestral plant, to trace the
CC   progeny of proprietary plants, in certification of a hybrid plant or to
CC   identify the progeny of a back-crossed plant with an ancestral plant.
XX
SQ   Sequence 41 BP; 12 A; 8 C; 6 G; 15 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 19;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 aaaaaattatttc 17
          |||||||||||||
Db     30 AAAAAATTATTTC 18


RESULT  26
AAT25074/c
ID   AAT25074 standard; DNA; 50 BP.
XX
AC   AAT25074;
XX
DT   22-OCT-1996  (first entry)
XX
DE   Human gene signature HUMGS07214.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 1763; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 50 BP; 26 A; 3 C; 3 G; 18 T; 0 other;


  Query Match             14.4%;  Score 13;  DB 16;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 tatttcaaagttt 24
          |||||||||||||
Db     43 TATTTCAAAGTTT 31


RESULT  27
AAX52184/c
ID   AAX52184 standard; DNA; 50 BP.
XX
AC   AAX52184;
```

XX
DT 18-JUN-1999 (first entry)
XX
DE Synthetic plasmid synlux4 construction oligonucleotide R68.
XX
KW DNA plasmid; lux A; lux B; Vibrio fisherii; luciferase; promoter;
KW tn9 kanamycin/neomycin phosphotransferase; DNA synthesis;
KW replication competent double-stranded polynucleotide; ss.
XX
OS Synthetic.
XX
PN WO9914318-A1.
XX
PD 25-MAR-1999.
XX
PF 16-SEP-1998; 98WO-US19312.
XX
PR 16-SEP-1997; 97US-0059017.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Evans GA;
XX
DR WPI; 1999-244029/20.
XX
PT Synthesis of replication competent double-stranded polynucleotides
XX
PS Example 4; Fig 5F; 135pp; English.
XX
CC AAX52021-212 represent oligonculeotide primers that were used to
CC construct a synthetic DNA plasmid sequence synlux4, to demonstrate the
CC method of the invention. Within the synlux4 sequence are included the
CC sequences of lux A, lux B, the A and B components of the Vibrio fisherii
CC luciferase sequence, positions of pUC19 including the origin of
CC replication and replication stability sequences, and the promoter and
CC coding sequence for tn9 kanamycin/neomycin phosphotransferase. The
CC specification describes a method for the synthesis of replication
CC competent double-stranded polynucleotides. The method comprises
CC generating a first set of oligonucleotides corresponding to the plus
CC strand and a second set corresponding to the minus strand and
CC annealing. The method can be used for preparing polynucleotides
CC encoding sequences involved in a biochemical pathway. In particular,
CC they can be used to produce polynucleotides encoding enzymes,
CC e.g. hexokinase, phosphohexose isomerase, phosphofructokinase-1,
CC aldolase, triose-phosphate isomerase, glyceraldehyde-3-phosphate
CC dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase,
CC enolase or pyruvate kinase. They can also be used for the preparation
CC of viral particles, artificial genomes and artificial genetic systems.
XX
SQ Sequence 50 BP; 19 A; 13 C; 7 G; 11 T; 0 other;

Query Match 14.4%; Score 13; DB 20; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 tattgttggagca 66
||||||||||||
Db 44 TATTGTTGGAGCA 32

RESULT 28
AAX52048
ID AAX52048 standard; DNA; 50 BP.
XX
AC AAX52048;
XX
DT 18-JUN-1999 (first entry)
XX
DE Synthetic plasmid synlux4 construction oligonucleotide F28.
XX
KW DNA plasmid; lux A; lux B; Vibrio fisherii; luciferase; promoter;
KW tn9 kanamycin/neomycin phosphotransferase; DNA synthesis;
KW replication competent double-stranded polynucleotide; ss.
XX
OS Synthetic.
XX
PN WO9914318-A1.
XX
PD 25-MAR-1999.
XX
PF 16-SEP-1998; 98WO-US19312.
XX
PR 16-SEP-1997; 97US-0059017.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Evans GA;
XX
DR WPI; 1999-244029/20.
XX
PT Synthesis of replication competent double-stranded polynucleotides
XX
PS Example 4; Fig 5A; 135pp; English.
XX
CC AAX52021-212 represent oligonculeotide primers that were used to
CC construct a synthetic DNA plasmid sequence synlux4, to demonstrate the
CC method of the invention. Within the synlux4 sequence are included the
CC sequences of lux A, lux B, the A and B components of the Vibrio fisherii
CC luciferase sequence, positions of pUC19 including the origin of
CC replication and replication stability sequences, and the promoter and
CC coding sequence for tn9 kanamycin/neomycin phosphotransferase. The
CC specification describes a method for the synthesis of replication
CC competent double-stranded polynucleotides. The method comprises
CC generating a first set of oligonucleotides corresponding to the plus
CC strand and a second set corresponding to the minus strand and
CC annealing. The method can be used for preparing polynucleotides
CC encoding sequences involved in a biochemical pathway. In particular,
CC they can be used to produce polynucleotides encoding enzymes,
CC e.g. hexokinase, phosphohexose isomerase, phosphofructokinase-1,
CC aldolase, triose-phosphate isomerase, glyceraldehyde-3-phosphate
CC dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase,
CC enolase or pyruvate kinase. They can also be used for the preparation
CC of viral particles, artificial genomes and artificial genetic systems.
XX
SQ Sequence 50 BP; 16 A; 7 C; 8 G; 19 T; 0 other;

Query Match 14.4%; Score 13; DB 20; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 tattgttggagca 66
||||||||||||
Db 33 tattgttggagca 45

RESULT 29
AAX70046/c
ID AAX70046 standard; RNA; 17 BP.
XX
AC AAX70046;
XX
DT 28-JUL-1999 (first entry)
XX
DE Human flt1 VEGF receptor hammerhead ribozyme substrate #1341.
XX
KW Vascular endothelial growth factor receptor; VEGF receptor; flt-1;
KW flk-1; KDR; hammerhead ribozyme; hairpin ribozyme; cleavage;
KW tumour angiogenesis; psoriasis; rheumatoid arthritis; ocular disease;
KW fms-like tyrosine kinase 1; kinase insert domain containing receptor;
KW foetal liver kinase 1; ss.
XX
OS Homo sapiens.
XX
PN WO9715662-A2.

```
XX
PD   01-MAY-1997.
XX
PF   25-OCT-1996;   96WO-US17480.
XX
PR   11-JAN-1996;   96US-0584040.
PR   26-OCT-1995;   95US-0005974.
XX
PA   (CHIR ) CHIRON CORP.
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Escobedo J,  McSwiggen J,  Pavco P,  Stinchcomb D;
XX
DR   WPI; 1997-259017/23.
XX
PT   Nucleic acid molecule modulating VEGF receptor(s) gene expression or
PT   mRNA stability  - useful for treating e.g. tumour angiogenesis,
PT   psoriasis, rheumatoid arthritis, etc., in a human patient
XX
PS   Claim 4; Page 87; 218pp; English.
XX
CC   The present invention describes nucleic acid molecules which modulate
CC   the synthesis, expression and/or stability of a mRNA encoding 1 or more
CC   receptors of vascular endothelial growth factor (VEGF). A patient
CC   (preferably human) having a condition associated with the level of the
CC   fms-like tyrosine kinase 1 (flt-1), kinase insert domain containing
CC   receptor (KDR) and/or foetal liver kinase 1 (flk-1) (e.g. tumour
CC   angiogenesis, ocular diseases, psoriasis and rheumatoid arthritis) can
CC   be treated by administering the nucleic acid molecule or the expression
CC   vector to the patient. AAX67275 to AAX75752 represent specific examples
CC   of nucleic acid molecules from the present invention.
XX
SQ   Sequence 17 BP; 8 A; 1 C; 4 G; 4 U; 0 other;


  Query Match             13.3%;  Score 12;  DB 18;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 atttcaaagttt 24
           ||||||||||||
Db      15 ATTTCAAAGTTT 4


RESULT  30
AAX70047/c
ID   AAX70047 standard; RNA; 17 BP.
XX
AC   AAX70047;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Human fltl VEGF receptor hammerhead ribozyme substrate #1342.
XX
KW   Vascular endothelial growth factor receptor; VEGF receptor; flt-1;
KW   flk-1; KDR; hammerhead ribozyme; hairpin ribozyme; cleavage;
KW   tumour angiogenesis; psoriasis; rheumatoid arthritis; ocular disease;
KW   fms-like tyrosine kinase 1; kinase insert domain containing receptor;
KW   foetal liver kinase 1; ss.
XX
OS   Homo sapiens.
XX
PN   WO9715662-A2.
XX
PD   01-MAY-1997.
XX
PF   25-OCT-1996;   96WO-US17480.
XX
PR   11-JAN-1996;   96US-0584040.
PR   26-OCT-1995;   95US-0005974.
XX
PA   (CHIR ) CHIRON CORP.
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Escobedo J,  McSwiggen J,  Pavco P,  Stinchcomb D;
XX
DR   WPI; 1997-259017/23.
XX
PT   Nucleic acid molecule modulating VEGF receptor(s) gene expression or
PT   mRNA stability  - useful for treating e.g. tumour angiogenesis,
PT   psoriasis, rheumatoid arthritis, etc., in a human patient
XX
PS   Claim 4; Page 87; 218pp; English.
XX
CC   The present invention describes nucleic acid molecules which modulate
CC   the synthesis, expression and/or stability of a mRNA encoding 1 or more
CC   receptors of vascular endothelial growth factor (VEGF). A patient
CC   (preferably human) having a condition associated with the level of the
CC   fms-like tyrosine kinase 1 (flt-1), kinase insert domain containing
CC   receptor (KDR) and/or foetal liver kinase 1 (flk-1) (e.g. tumour
CC   angiogenesis, ocular diseases, psoriasis and rheumatoid arthritis) can
CC   be treated by administering the nucleic acid molecule or the expression
CC   vector to the patient. AAX67275 to AAX75752 represent specific examples
CC   of nucleic acid molecules from the present invention.
XX
SQ   Sequence 17 BP; 8 A; 1 C; 3 G; 5 U; 0 other;


  Query Match             13.3%;  Score 12;  DB 18;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 atttcaaagttt 24
           ||||||||||||
Db      14 ATTTCAAAGTTT 3


RESULT  31
AAA21453
ID   AAA21453 standard; RNA; 17 BP.
XX
AC   AAA21453;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Integrin alpha 6 subunit substrate sequence SEQ ID NO:4679.
XX
KW   Human; aryl hydrocarbon nuclear transport; ARNT; TIE-2; angiogenesis;
KW   integrin alpha 6 subunit; integrin subunit beta 3; hairpin ribozyme;
KW   hammerhead ribozyme; angiogenic factor; cytostatic; antidiabetic;
KW   ophthalmologic; antiinflammatory; antiarthritic; antipsoriatic; ARMD;
KW   dermatological; RNA cleavage; cancer; diabetic retinopathy; arthritis;
KW   age related macular degeneration; inflammation; neovascular glaucoma;
KW   myopic degeneration; psoriasis; verruca vulgaris; angiofibroma;
KW   tuberous sclerosis; pot-wine stain; Sturge Weber syndrome;
KW   Kippel-Trenaunay-Weber syndrome; Osler-Weber-Rendu syndrome; ss.
XX
OS   Homo sapiens.
XX
PN   WO9950403-A2.
XX
PD   07-OCT-1999.
XX
PF   24-MAR-1999;   99WO-US06507.
XX
PR   27-MAR-1998;   98US-0079678.
XX
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Pavco PA,  Roberts E,  Jarvis T,  Coeshott C,  McSwiggen JA;
XX
DR   WPI; 1999-591315/50.
XX
PT   Novel ribozymes for modulating the synthesis, expression and/or
PT   stability of an mRNA encoding an angiogenic factors  -
XX
```

```
PS   Claim 55; Page 209; 305pp; English.
XX
CC   The present invention describes enzymatic nucleic acid molecules with
CC   RNA cleaving activity, which specifically cleave RNA encoded by an aryl
CC   hydrocarbon nuclear transporter (ARNT) gene, an integrin subunit beta 3
CC   gene, an integrin alpha 6 subunit gene, or a Tie-2 gene. AAA16775 to
CC   AAA17167 and AAA17561 to AAA17622 represent ribozyme sequences for ARNT,
CC   and AAA17168 to AAA17560 and AAA17623 to AAA17684 represent their
CC   corresponding target sequences; AAA17685 to AAA18385 and AAA19087 to
CC   AAA19154 represent ribozyme sequences for Tie-2, and AAA18386 to AAA19086
CC   and AAA19155 to AAA19222 represent their corresponding target sequences;
CC   AAA19223 to AAA20361 and AAA21501 to AAA21595 represent ribozyme
CC   sequences for integrin alpha 6 subunit, and AAA20362 to AAA21500 and
CC   AAA21596 to AAA21688 represent their corresponding target sequences;
CC   AAA21689 to AAA22475 and AAA23263 to AAA23342 represent ribozyme sequence
CC   for integrin subunit beta 3, and AAA22476 to AAA23262, AAA23343 to
CC   AAA23422 represent their corresponding target sequences. The ribozymes of
CC   the invention are used for modulating the synthesis, expression and/or
CC   stability of an mRNA encoding angiogenic factor, especially ARNT,
CC   integrin subunit beta-3, integrin subunit alpha-6, or Tie-2. They are
CC   especially used to treat cancer, diabetic retinopathy, age related
CC   macular degeneration (ARMD), inflammation, and arthritis, as well as
CC   neovascular glaucoma, myopic degeneration, psoriasis, verruca vulgaris,
CC   angiofibroma of tuberous sclerosis, pot-wine stains, Sturge Weber
CC   syndrome, Kippel-Trenaunay-Weber syndrome, Osler-Weber-Rendu syndrome,
CC   and other syndromes and diseases related to the levels of ARNT, Tie-2,
CC   integrin subunit alpha-6, or integrin subunit beta-3.
XX
SQ   Sequence 17 BP; 6 A; 1 C; 2 G; 8 U; 0 other;

  Query Match             13.3%;  Score 12;  DB 20;  Length 17;
  Best Local Similarity   58.3%;  Pred. No. 8.3e+03;
  Matches    7;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 aaaattatttca 18
           ||||::|:::||
Db       4 aaaauuauuuca 15


RESULT  32
AAA21454
ID   AAA21454 standard; RNA; 17 BP.
XX
AC   AAA21454;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Integrin alpha 6 subunit substrate sequence SEQ ID NO:4680.
XX
KW   Human; aryl hydrocarbon nuclear transport; ARNT; TIE-2; angiogenesis;
KW   integrin alpha 6 subunit; integrin subunit beta 3; hairpin ribozyme;
KW   hammerhead ribozyme; angiogenic factor; cytostatic; antidiabetic;
KW   ophthalmologic; antiinflammatory; antiarthritic; antipsoriatic; ARMD;
KW   dermatological; RNA cleavage; cancer; diabetic retinopathy; arthritis;
KW   age related macular degeneration; inflammation; neovascular glaucoma;
KW   myopic degeneration; psoriasis; verruca vulgaris; angiofibroma;
KW   tuberous sclerosis; pot-wine stain; Sturge Weber syndrome;
KW   Kippel-Trenaunay-Weber syndrome; Osler-Weber-Rendu syndrome; ss.
XX
OS   Homo sapiens.
XX
PN   WO9950403-A2.
XX
PD   07-OCT-1999.
XX
PF   24-MAR-1999;   99WO-US06507.
XX
PR   27-MAR-1998;   98US-0079678.
XX
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Pavco PA,  Roberts E,  Jarvis T,  Coeshott C,  McSwiggen JA;
XX
DR   WPI; 1999-591315/50.
XX
PT   Novel ribozymes for modulating the synthesis, expression and/or
PT   stability of an mRNA encoding an angiogenic factors  -
XX
PS   Claim 55; Page 209; 305pp; English.
XX
CC   The present invention describes enzymatic nucleic acid molecules with
CC   RNA cleaving activity, which specifically cleave RNA encoded by an aryl
CC   hydrocarbon nuclear transporter (ARNT) gene, an integrin subunit beta 3
CC   gene, an integrin alpha 6 subunit gene, or a Tie-2 gene. AAA16775 to
CC   AAA17167 and AAA17561 to AAA17622 represent ribozyme sequences for ARNT,
CC   and AAA17168 to AAA17560 and AAA17623 to AAA17684 represent their
CC   corresponding target sequences; AAA17685 to AAA18385 and AAA19087 to
CC   AAA19154 represent ribozyme sequences for Tie-2, and AAA18386 to AAA19086
CC   and AAA19155 to AAA19222 represent their corresponding target sequences;
CC   AAA19223 to AAA20361 and AAA21501 to AAA21595 represent ribozyme
CC   sequences for integrin alpha 6 subunit, and AAA20362 to AAA21500 and
CC   AAA21596 to AAA21688 represent their corresponding target sequences;
CC   AAA21689 to AAA22475 and AAA23263 to AAA23342 represent ribozyme sequence
CC   for integrin subunit beta 3, and AAA22476 to AAA23262, AAA23343 to
CC   AAA23422 represent their corresponding target sequences. The ribozymes of
CC   the invention are used for modulating the synthesis, expression and/or
CC   stability of an mRNA encoding angiogenic factor, especially ARNT,
CC   integrin subunit beta-3, integrin subunit alpha-6, or Tie-2. They are
CC   especially used to treat cancer, diabetic retinopathy, age related
CC   macular degeneration (ARMD), inflammation, and arthritis, as well as
CC   neovascular glaucoma, myopic degeneration, psoriasis, verruca vulgaris,
CC   angiofibroma of tuberous sclerosis, pot-wine stains, Sturge Weber
CC   syndrome, Kippel-Trenaunay-Weber syndrome, Osler-Weber-Rendu syndrome,
CC   and other syndromes and diseases related to the levels of ARNT, Tie-2,
CC   integrin subunit alpha-6, or integrin subunit beta-3.
XX
SQ   Sequence 17 BP; 7 A; 1 C; 2 G; 7 U; 0 other;

  Query Match             13.3%;  Score 12;  DB 20;  Length 17;
  Best Local Similarity   58.3%;  Pred. No. 8.3e+03;
  Matches    7;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 aaaattatttca 18
           ||||::|:::||
Db       3 aaaauuauuuca 14


RESULT  33
AAA21455
ID   AAA21455 standard; RNA; 17 BP.
XX
AC   AAA21455;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Integrin alpha 6 subunit substrate sequence SEQ ID NO:4681.
XX
KW   Human; aryl hydrocarbon nuclear transport; ARNT; TIE-2; angiogenesis;
KW   integrin alpha 6 subunit; integrin subunit beta 3; hairpin ribozyme;
KW   hammerhead ribozyme; angiogenic factor; cytostatic; antidiabetic;
KW   ophthalmologic; antiinflammatory; antiarthritic; antipsoriatic; ARMD;
KW   dermatological; RNA cleavage; cancer; diabetic retinopathy; arthritis;
KW   age related macular degeneration; inflammation; neovascular glaucoma;
KW   myopic degeneration; psoriasis; verruca vulgaris; angiofibroma;
KW   tuberous sclerosis; pot-wine stain; Sturge Weber syndrome;
KW   Kippel-Trenaunay-Weber syndrome; Osler-Weber-Rendu syndrome; ss.
XX
OS   Homo sapiens.
XX
PN   WO9950403-A2.
XX
PD   07-OCT-1999.
XX
PF   24-MAR-1999;   99WO-US06507.
```

XX
PR 27-MAR-1998; 98US-0079678.
XX
PA (RIBO-) RIBOZYME PHARM INC.
XX
PI Pavco PA, Roberts E, Jarvis T, Coeshott C, McSwiggen JA;
XX
DR WPI; 1999-591315/50.
XX
PT Novel ribozymes for modulating the synthesis, expression and/or
PT stability of an mRNA encoding an angiogenic factors -
XX
PS Claim 55; Page 209; 305pp; English.
XX
CC The present invention describes enzymatic nucleic acid molecules with
CC RNA cleaving activity, which specifically cleave RNA encoded by an aryl
CC hydrocarbon nuclear transporter (ARNT) gene, an integrin subunit beta 3
CC gene, an integrin alpha 6 subunit gene, or a Tie-2 gene. AAA16775 to
CC AAA17167 and AAA17561 to AAA17622 represent ribozyme sequences for ARNT,
CC and AAA17168 to AAA17560 and AAA17623 to AAA17684 represent their
CC corresponding target sequences; AAA17685 to AAA18385 and AAA19087 to
CC AAA19154 represent ribozyme sequences for Tie-2, and AAA18386 to AAA19086
CC and AAA19155 to AAA19222 represent their corresponding target sequences;
CC AAA19223 to AAA20361 and AAA21501 to AAA21595 represent ribozyme
CC sequences for integrin alpha 6 subunit, and AAA20362 to AAA21500 and
CC AAA21596 to AAA21688 represent their corresponding target sequences;
CC AAA21689 to AAA22475 and AAA23263 to AAA23342 represent ribozyme sequence
CC for integrin subunit beta 3, and AAA22476 to AAA23262, AAA23343 to
CC AAA23422 represent their corresponding target sequences. The ribozymes of
CC the invention are used for modulating the synthesis, expression and/or
CC stability of an mRNA encoding angiogenic factor, especially ARNT,
CC integrin subunit beta-3, integrin subunit alpha-6, or Tie-2. They are
CC especially used to treat cancer, diabetic retinopathy, age related
CC macular degeneration (ARMD), inflammation, and arthritis, as well as
CC neovascular glaucoma, myopic degeneration, psoriasis, verruca vulgaris,
CC angiofibroma of tuberous sclerosis, pot-wine stains, Sturge Weber
CC syndrome, Kippel-Trenaunay-Weber syndrome, Osler-Weber-Rendu syndrome,
CC and other syndromes and diseases related to the levels of ARNT, Tie-2,
CC integrin subunit alpha-6, or integrin subunit beta-3.
XX
SQ Sequence 17 BP; 7 A; 1 C; 1 G; 8 U; 0 other;

Query Match 13.3%; Score 12; DB 20; Length 17;
Best Local Similarity 58.3%; Pred. No. 8.3e+03;
Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 7 aaaattatttca 18
|||| :| ::: ||
Db 1 aaaauuauuuca 12

RESULT 34
AAF03088/c
ID AAF03088 standard; DNA; 17 BP.
XX
AC AAF03088;
XX
DT 16-FEB-2001 (first entry)
XX
DE Hammerhead ribozyme substrate #1383.
XX
KW Ribozyme; erythropoietin; granulocyte colony stimulating factor;
KW interferon alpha; ss.
XX
OS Homo sapiens.
XX
PN WO200061729-A2.
XX
PD 19-OCT-2000.
XX
PF 11-APR-2000; 2000WO-US09721.
XX
PR 12-APR-1999; 99US-0129390.
XX
PA (RIBO-) RIBOZYME PHARM INC.
XX
PI Blatt L, Zwick M, Pavco P, McSwiggen J;
XX
DR WPI; 2000-647423/62.
XX
PT Enzymatic and antisense nucleic acid inhibition of repressor genes,
PT useful for producing e.g. granulocyte colony stimulating factor
PT protein, interferon alpha and erythropoietin -
XX
PS Claim 37; Page 87; 164pp; English.
XX
CC The present invention relates to enzymatic and antisense nucleic acid
CC molecules that act as inhibitors of the expression of repressor genes
CC encoding the TR2 Orphan receptor, EAR3/COUP-TF-1, the GATA
CC transcription factor gene, IRF-2 and/or the CAATT Displacement
CC Protein (CDP). Inhibition of the repressors removes prevents
CC inhibition (and consequently increases expression of) genes involved in
CC the production of erythropoietin, granulocyte colony stimulating factor
CC protein and interferon alpha.
XX
SQ Sequence 17 BP; 6 A; 1 C; 0 G; 10 T; 0 other;

Query Match 13.3%; Score 12; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aaaaaaattatt 15
||||||||||||
Db 17 AAAAAAATTATT 6

RESULT 35
AAF03089/c
ID AAF03089 standard; DNA; 17 BP.
XX
AC AAF03089;
XX
DT 16-FEB-2001 (first entry)
XX
DE Hammerhead ribozyme substrate #1384.
XX
KW Ribozyme; erythropoietin; granulocyte colony stimulating factor;
KW interferon alpha; ss.
XX
OS Homo sapiens.
XX
PN WO200061729-A2.
XX
PD 19-OCT-2000.
XX
PF 11-APR-2000; 2000WO-US09721.
XX
PR 12-APR-1999; 99US-0129390.
XX
PA (RIBO-) RIBOZYME PHARM INC.
XX
PI Blatt L, Zwick M, Pavco P, McSwiggen J;
XX
DR WPI; 2000-647423/62.
XX
PT Enzymatic and antisense nucleic acid inhibition of repressor genes,
PT useful for producing e.g. granulocyte colony stimulating factor
PT protein, interferon alpha and erythropoietin -
XX
PS Claim 37; Page 87; 164pp; English.
XX
CC The present invention relates to enzymatic and antisense nucleic acid
CC molecules that act as inhibitors of the expression of repressor genes
CC encoding the TR2 Orphan receptor, EAR3/COUP-TF-1, the GATA

```
CC   transcription factor gene, IRF-2 and/or the CAATT Displacement
CC   Protein (CDP). Inhibition of the repressors removes prevents
CC   inhibition (and consequently increases expression of) genes involved in
CC   the production of erythropoietin, granulocyte colony stimulating factor
CC   protein and interferon alpha.
XX
SQ   Sequence 17 BP; 8 A; 1 C; 0 G; 8 T; 0 other;


  Query Match           13.3%;  Score 12;  DB 21;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatt 15
          ||||||||||||
Db     14 AAAAAAATTATT 3


RESULT  36
AAF03090/c
ID   AAF03090 standard; DNA; 17 BP.
XX
AC   AAF03090;
XX
DT   16-FEB-2001  (first entry)
XX
DE   Hammerhead ribozyme substrate #1385.
XX
KW   Ribozyme; erythropoietin; granulocyte colony stimulating factor;
KW   interferon alpha; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061729-A2.
XX
PD   19-OCT-2000.
XX
PF   11-APR-2000; 2000WO-US09721.
XX
PR   12-APR-1999;   99US-0129390.
XX
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Blatt L,  Zwick M,  Pavco P,  McSwiggen J;
XX
DR   WPI; 2000-647423/62.
XX
PT   Enzymatic and antisense nucleic acid inhibition of repressor genes,
PT   useful for producing e.g. granulocyte colony stimulating factor
PT   protein, interferon alpha and erythropoietin -
XX
PS   Claim 37; Page 87; 164pp; English.
XX
CC   The present invention relates to enzymatic and antisense nucleic acid
CC   molecules that act as inhibitors of the expression of repressor genes
CC   encoding the TR2 Orphan receptor, EAR3/COUP-TF-1, the GATA
CC   transcription factor gene, IRF-2 and/or the CAATT Displacement
CC   Protein (CDP). Inhibition of the repressors removes prevents
CC   inhibition (and consequently increases expression of) genes involved in
CC   the production of erythropoietin, granulocyte colony stimulating factor
CC   protein and interferon alpha.
XX
SQ   Sequence 17 BP; 8 A; 1 C; 0 G; 8 T; 0 other;


  Query Match           13.3%;  Score 12;  DB 21;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatt 15
          ||||||||||||
Db     13 AAAAAAATTATT 2


RESULT  37
AAF03091/c
ID   AAF03091 standard; DNA; 17 BP.
XX
AC   AAF03091;
XX
DT   16-FEB-2001  (first entry)
XX
DE   Hammerhead ribozyme substrate #1386.
XX
KW   Ribozyme; erythropoietin; granulocyte colony stimulating factor;
KW   interferon alpha; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061729-A2.
XX
PD   19-OCT-2000.
XX
PF   11-APR-2000; 2000WO-US09721.
XX
PR   12-APR-1999;   99US-0129390.
XX
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Blatt L,  Zwick M,  Pavco P,  McSwiggen J;
XX
DR   WPI; 2000-647423/62.
XX
PT   Enzymatic and antisense nucleic acid inhibition of repressor genes,
PT   useful for producing e.g. granulocyte colony stimulating factor
PT   protein, interferon alpha and erythropoietin -
XX
PS   Claim 37; Page 87; 164pp; English.
XX
CC   The present invention relates to enzymatic and antisense nucleic acid
CC   molecules that act as inhibitors of the expression of repressor genes
CC   encoding the TR2 Orphan receptor, EAR3/COUP-TF-1, the GATA
CC   transcription factor gene, IRF-2 and/or the CAATT Displacement
CC   Protein (CDP). Inhibition of the repressors removes prevents
CC   inhibition (and consequently increases expression of) genes involved in
CC   the production of erythropoietin, granulocyte colony stimulating factor
CC   protein and interferon alpha.
XX
SQ   Sequence 17 BP; 9 A; 0 C; 0 G; 8 T; 0 other;


  Query Match           13.3%;  Score 12;  DB 21;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatt 15
          ||||||||||||
Db     12 AAAAAAATTATT 1


RESULT  38
AAA08921
ID   AAA08921 standard; DNA; 18 BP.
XX
AC   AAA08921;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human survivin DNA antisense oligonucleotide, ISIS 23663.
XX
KW   Survivin; inhibitor of apoptosis; IAP; caspase inhibitor; caspase-3;
KW   cell cycle regulation; cancer; cytostatic; antisense oligonucleotide; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
```

```
FH   Key             Location/Qualifiers
FT   modified_base   1..18
FT                   /*tag=  a
FT                   /note= "phosphorothioate backbone"
XX
PN   WO200018781-A1.
XX
PD   06-APR-2000.
XX
PF   23-SEP-1999;   99WO-US22076.
XX
PR   29-SEP-1998;   98US-0163162.
PR   05-APR-1999;   99US-0286407.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Ackermann EJ,  Swayze EE,  Cowsert LM;
XX
DR   WPI; 2000-293103/25.
XX
PT   Antisense molecules targeted to Survivin, useful for inducing apoptosis
PT   in cancer cells
XX
PS   Claim 3; Page 66; 73pp; English.
XX
CC   This is an antisense oligonucleotide targeted to the coding sequence,
CC   nucleotide 393, of human survivin DNA (see AAA08903).  AAA08910-49 were
CC   analyzed for effect on survivin mRNA levels by quantitative real-time
CC   PCR.  The data obtained were averages from three experiments.  This
CC   antisense oligonucleotide provided 0% inhibition of surivin mRNA.  It
CC   was found that ISIS 23667 (AAA08925) provided 70% inhibition and ISIS
CC   23672 (AAA08930) provided 64% inhibition.
CC   Survivin, an IAP (inhibitor of apoptosis) Caspase inhibitor,
CC   has been found to be involved in cell cycle regulation and is expressed
CC   in the G2/M phase of the cell-cycle in a cell cycle regulated manner and
CC   associates with microtubules of the mitotic spindle.  Disruption of this
CC   interaction results in loss of survivin's anti-apoptotic function and
CC   increased caspase-3 activity during mitosis.  Caspase-3 is associated
CC   with apoptotic cell death.  It is therefore believed that survivin may
CC   counteract a default induction of apoptosis in the G2/M phase.  It is
CC   also believed that the over expression of survivin in cancer may
CC   overcome this apoptotic check point, allowing undesired survival and
CC   division of cancer cells.  Antisense oligonucleotides (ASO's) may be used
CC   to down regulate endogenous survivin and to increase caspase-3-dependent
CC   apoptosis in cells in the G2/M phase.
XX
SQ   Sequence 18 BP; 1 A; 3 C; 3 G; 11 T; 0 other;


  Query Match             13.3%;  Score 12;  DB 21;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      51 tcttattgttgg 62
           ||||||||||||
Db       1 tcttattgttgg 12


RESULT  39
AAQ62806/c
ID   AAQ62806 standard; DNA; 19 BP.
XX
AC   AAQ62806;
XX
DT   21-JUL-1994  (first entry)
XX
DE   Env gene 5' primer.
XX
KW   FIV; feline immunodeficiency virus; primer; env; cat; petaluma;
KW   amplification; polymerase chain reaction; PCR; vaccine; ss.
XX
OS   Synthetic.
XX
PN   WO9402613-A.
XX
PD   03-FEB-1994.
XX
PF   15-JUL-1993;   93WO-EP01861.
XX
PR   17-JUL-1992;   92GB-0015233.
XX
PA   (PITM ) PITMAN MOORE INC.
XX
PI   Osterhaus ADME,  Rimmelzwaan GF,  Siebelink CHJ;
XX
DR   WPI; 1994-048871/06.
XX
PT   Synthetic feline immunodeficiency virus (FIV) polypeptide - for
PT   use in vaccines used to combat FIV
XX
PS   Example 1; Page 21; 41pp; English.
XX
CC   The FIV env gene was amplified by PCR using bone marrow
CC   derived DNA, obtained from a FIV-infected cat as source
CC   of FIV proviral DNA.  The primers used are given in AAQ62805-06
CC   and AAQ62808-09 and were derived from the nucleotide sequence
CC   of the petaluma strain of FIV.
XX
SQ   Sequence 19 BP; 9 A; 3 C; 5 G; 2 T; 0 other;


  Query Match             13.3%;  Score 12;  DB 15;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      50 ttcttattgttg 61
           ||||||||||||
Db      13 TTCTTATTGTTG 2


RESULT  40
AAZ69625/c
ID   AAZ69625 standard; DNA; 19 BP.
XX
AC   AAZ69625;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human biallelic marker upstream amplification primer SEQ ID NO:3981.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   amplification; single nucleotide polymorphism; SNP; PCR primer;
KW   diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
```

PS Claim 8; Page 1080; 2745pp; English.
XX
CC AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC invention, which contain a polymorphic base at position 24 of their
CC nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC primers for the biallelic markers. The biallelic markers of the
CC invention have a variety of uses: they can be used for high density
CC mapping of the human genome, and in complex association studies and
CC haplotyping studies which are useful in determining the genetic basis
CC for disease states. Compositions and methods of the invention can also
CC be useful for the identification of the targets for the development of
CC pharmaceutical agents and diagnostic methods, as well as the
CC characterisation of the differential efficacious responses to and side
CC effects from pharmaceutical agents acting on a disease as well as other
CC treatment.
CC N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC and 3367, are not actually given a sequence in the Sequence Listing
CC from the present invention.
XX
SQ Sequence 19 BP; 1 A; 4 C; 5 G; 9 T; 0 other;

Query Match 13.3%; Score 12; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 ccaatgaaagca 84
||||||||||||
Db 18 CCAATGAAAGCA 7

RESULT 41
AAZ75126/c
ID AAZ75126 standard; DNA; 19 BP.
XX
AC AAZ75126;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human biallelic marker downstream amplification primer SEQ ID NO:9482.
XX
KW Human genome; biallelic marker; high density disequilibrium map;
KW genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW haplotyping; hybridisation; identification; characterisation;
KW amplification; single nucleotide polymorphism; SNP; PCR primer;
KW diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9954500-A2.
XX
PD 28-OCT-1999.
XX
PF 21-APR-1999; 99WO-IB00822.
XX
PR 21-APR-1998; 98US-0082614.
PR 23-NOV-1998; 98US-0109732.
XX
PA (GEST) GENSET.
XX
PI Cohen D, Blumenfeld M, Chumakov I;
XX
DR WPI; 2000-013267/01.
XX
PT Novel biallelic markers used to construct a high density disequilibrium
PT map of the human genome -
XX
PS Claim 8; Page 2252; 2745pp; English.
XX
CC AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC invention, which contain a polymorphic base at position 24 of their
CC nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC primers for the biallelic markers. The biallelic markers of the
CC invention have a variety of uses: they can be used for high density
CC mapping of the human genome, and in complex association studies and
CC haplotyping studies which are useful in determining the genetic basis
CC for disease states. Compositions and methods of the invention can also
CC be useful for the identification of the targets for the development of
CC pharmaceutical agents and diagnostic methods, as well as the
CC characterisation of the differential efficacious responses to and side
CC effects from pharmaceutical agents acting on a disease as well as other
CC treatment.
CC N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC and 3367, are not actually given a sequence in the Sequence Listing
CC from the present invention.
XX
SQ Sequence 19 BP; 7 A; 3 C; 4 G; 5 T; 0 other;

Query Match 13.3%; Score 12; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 gtttactgtaat 32
||||||||||||
Db 16 GTTTACTGTAAT 5

RESULT 42
AAQ05907
ID AAQ05907 standard; DNA; 20 BP.
XX
AC AAQ05907;
XX
DT 10-JAN-1991 (first entry)
XX
DE HIV mRNA translation inhibiting antisense oligonucleotide (d).
XX
KW Modified antisense oligonucleotide; HIV mRNA translation inhibition;
KW HIV-1; transactivator region; leader sequence; tat gene; probe; ss.
XX
OS Synthetic.
XX
PN EP386563-A.
XX
PD 12-SEP-1990.
XX
PF 25-FEB-1990; 90EP-0103641.
XX
PR 09-MAR-1989; 89DE-3907562.
XX
PA (FARB) BAYER AG.
XX
PI Stropp U, Baumgarten J, Lobberding A, Springer W, Piel N;
PI Kretschmer, Kolbl H, Frommer W;
XX
DR WPI; 1990-276634/37.
XX
PT Chemically modified antisense oligo-nucleotide(s) - inhibiting
PT translation of HIV MRNA
XX
PS Claim 3(d); Page 8; 15pp; German.
XX
CC The chemically modified sequences are opposed to the HIV-1
CC transactivator region, leader sequences (nt 21-53, 74-161, 202-279)
CC or exon 2 or 3 of the tat gene (nt 5369-5403, 5421-5548, 5583-5617,
CC 7967-8366, 8385-9183). This sequence comprises nt 5493-5512
CC and gives 85% inhibition of HIV-1 tat translation in vitro
CC at concns. of 5-25 microM.
CC The sequence is useful for treating HIV infections and their
CC complementary sequences can be used as probes to detect HIV.
CC Chemical modification comprises replacing one or more internt
CC phosphodiester linkages by phosphorothioate or methylphosphonate
CC linkages.
CC See also AAQ05904-11.
XX

```
SQ   Sequence 20 BP; 8 A; 4 C; 3 G; 5 T; 0 other;

  Query Match           13.3%;  Score 12;  DB 11;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 caatgaaagcaa 13


RESULT  43
AAT80813/c
ID   AAT80813 standard; cDNA; 20 BP.
XX
AC   AAT80813;
XX
DT   14-APR-1998  (first entry)
XX
DE   Staphylococcus aureus Gene #22 PCR primer sequence 4.
XX
KW   Staphylococcus aureus WCUH 29; antagonist; antibacterial; immunogen;
KW   vaccine; disease; protection; isolation; PCR primer; ss.
XX
OS   Synthetic.
OS   Staphylococcus aureus.
XX
PN   WO9731114-A2.
XX
PD   28-AUG-1997.
XX
PF   25-FEB-1997;   97WO-GB00524.
XX
PR   26-FEB-1996;   96GB-0004045.
XX
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Burnham MKR,  Hodgson JE;
XX
DR   WPI; 1997-435166/40.
XX
PT   New Staphylococcus aureus polynucleotide and polypeptide(s) - for
PT   isolating antagonist of the polypeptide(s) useful as anti-bacterials
XX
PS   Disclosure; Page 52; 117pp; English.
XX
CC   The present sequence represents a PCR primer used in the present
CC   invention describing novel polypeptides, which can optionally be
CC   expressed in NCIMB 40771. The polypeptides, and polynucleotides encoding
CC   it, are derived from Staphylococcus aureus. Cells expressing ligands
CC   binding the polypeptides can be used to isolated candidate compounds
CC   that bind and inhibit the activity of the polypeptides. Such compounds
CC   can be used as anti-bacterial compounds. The polypeptides may also be
CC   used as immunogens to vaccinate an animal for protection against
CC   Staphylococcus aureus caused disease.
XX
SQ   Sequence 20 BP; 11 A; 4 C; 2 G; 3 T; 0 other;

  Query Match           13.3%;  Score 12;  DB 18;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      20 GTTTTCATGTTT 9


RESULT  44
AAV72697/c
ID   AAV72697 standard; DNA; 20 BP.
XX
AC   AAV72697;
XX
DT   17-FEB-1999  (first entry)
XX
DE   Corn kernel oil concentration controlling loci marker s1384 primer 2.
XX
KW   Corn; kernel oil; concentration; trait controlling loci; genetic marker;
KW   Zea mays; breeding; PCR primer; ss.
XX
OS   Synthetic.
OS   Zea mays.
XX
PN   WO9842870-A1.
XX
PD   01-OCT-1998.
XX
PF   19-MAR-1998;   98WO-US05550.
XX
PR   24-MAR-1997;   97US-0041515.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Reiter RS;
XX
DR   WPI; 1998-609896/51.
XX
PT   Breeding corn with increased oil concentration - comprises use of
PT   genetic markers to identify trait loci controlling kernel oil
PT   concentration
XX
PS   Example 2; Page 5; 50pp; English.
XX
CC   A new method has been developed of breeding for corn with increased
CC   kernel oil concentration. The method comprises: (a) selecting a corn
CC   plant from a breeding population using at least one of the genetic
CC   markers s1375, s1384, s1394, s1416, s1422, s1432, s1457, s1480, s1476,
CC   s1478, s1484, s1500, s1513, s1529, s1544, s1545, s1630, s1633, s1647,
CC   s1750, s1756, s1757, s1767, s1772, s1774, s1780, s1797, s1813, s1816,
CC   s1817, s1836, s1853, s1860, s1870, s1921, s1922, s1925, s1931, s1933,
CC   s1939, s1946, s1949, s2054, s2055, s2057, s2058, s2097, s2122, s2125,
CC   2150, s2156, and s2175; and (b) crossing the selected plant with a
CC   second plant and obtaining progeny with increased kernel oil
CC   concentration. Also described are: (1) a method for identifying corn
CC   plants or lines for use a parents to create a breeding population,
CC   comprising: (a) genotyping corn plants or lines with one or more of the
CC   above genetic markers; and (b) identifying plants or lines which are
CC   predicted to produce transgressive segregants for kernel oil
CC   concentration; and (2) trait loci controlling kernel oil concentration
CC   mapped by the above genetic markers, with the exception of s1480.
CC   AAV72694 to AAV72797 represent PCR primers which are used to amplify the
CC   genetic markers for use in the method of the invention.
XX
SQ   Sequence 20 BP; 4 A; 7 C; 3 G; 6 T; 0 other;

  Query Match           13.3%;  Score 12;  DB 19;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      77 tgaaagcaagtg 88
           ||||||||||||
Db      12 TGAAAGCAAGTG 1


RESULT  45
AAX92327
ID   AAX92327 standard; DNA; 20 BP.
XX
AC   AAX92327;
XX
DT   13-SEP-1999  (first entry)
XX
DE   PCR primer used to amplify an ORF of Chlamydia pneumoniae.
```

XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; PCR primer; ss.
XX
OS Synthetic.
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1503; Disclosure; 1912pp; English.
XX
CC AAX91991-X97517 represent PCR primers used to amplify open reading
CC frames and other nucleic acid sequences from the genome of
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
CC disease such as pneumonia and bronchitis and is thought to be a
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC containing C. pneumoniae nucleotides sequences can also be used as
CC immunogenic compositions, especially where the vector directs the
CC expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 20 BP; 3 A; 5 C; 5 G; 7 T; 0 other;

Query Match 13.3%; Score 12; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 atgttttcatgt 48
||||||||||||
Db 9 atgttttcatgt 20

Search completed: January 24, 2002, 03:28:21
Job time: 3673 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:22:33 ; Search time 93.51 Seconds
(without alignments)
791.983 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 495388

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16	4.9	29	3	US-08-816-977-12	Sequence 12, Appl
	2	16	4.9	36	1	US-08-629-600-16	Sequence 16, Appl
	3	15	4.6	27	1	US-08-120-827-66	Sequence 66, Appl
	4	15	4.6	27	1	US-08-478-675-66	Sequence 66, Appl
	5	15	4.6	30	5	PCT-US92-10792-3	Sequence 3, Appli
c	6	15	4.6	32	1	US-08-256-261-29	Sequence 29, Appl
c	7	15	4.6	32	3	US-08-852-299-29	Sequence 29, Appl
c	8	14	4.3	18	2	US-09-205-204-20	Sequence 20, Appl
c	9	14	4.3	29	3	US-08-816-977-12	Sequence 12, Appl
	10	14	4.3	30	2	US-08-629-001A-31	Sequence 31, Appl
	11	14	4.3	30	4	US-08-642-274D-110	Sequence 110, App
c	12	14	4.3	36	1	US-08-629-600-16	Sequence 16, Appl
	13	14	4.3	37	2	US-08-403-853-8	Sequence 8, Appli
	14	13	4.0	18	3	US-08-847-844A-113	Sequence 113, App
	15	13	4.0	18	4	US-08-686-968C-13	Sequence 13, Appl
	16	13	4.0	20	3	US-09-288-461-79	Sequence 79, Appl
c	17	13	4.0	21	3	US-08-691-045-61	Sequence 61, Appl
c	18	13	4.0	24	3	US-08-672-215-1	Sequence 1, Appli
	19	13	4.0	28	1	US-08-120-827-64	Sequence 64, Appl
	20	13	4.0	28	1	US-08-478-675-64	Sequence 64, Appl
	21	13	4.0	30	2	US-08-629-001A-79	Sequence 79, Appl
	22	13	4.0	30	4	US-08-642-274D-158	Sequence 158, App
c	23	13	4.0	31	1	US-08-330-638D-5	Sequence 5, Appli
c	24	13	4.0	31	2	US-08-906-746A-5	Sequence 5, Appli
c	25	13	4.0	36	1	US-08-247-809A-14	Sequence 14, Appl
c	26	13	4.0	36	2	US-08-711-728-14	Sequence 14, Appl
	27	13	4.0	37	2	US-08-097-554A-45	Sequence 45, Appl
	28	13	4.0	37	2	US-08-484-575A-10	Sequence 10, Appl
	29	13	4.0	37	3	US-08-477-459-10	Sequence 10, Appl
	30	13	4.0	37	3	US-08-480-640A-45	Sequence 45, Appl
	31	13	4.0	37	3	US-08-479-869-10	Sequence 10, Appl
	32	13	4.0	37	3	US-08-295-802-45	Sequence 45, Appl
	33	13	4.0	37	3	US-08-486-414-10	Sequence 10, Appl
	34	13	4.0	37	4	US-08-488-237A-45	Sequence 45, Appl
	35	13	4.0	37	5	PCT-US94-01826A-10	Sequence 10, Appl
	36	13	4.0	37	5	PCT-US94-02252A-10	Sequence 10, Appl
c	37	13	4.0	38	5	PCT-US96-00547-40	Sequence 40, Appl
c	38	13	4.0	39	1	US-08-105-483-168	Sequence 168, App
c	39	13	4.0	39	1	US-08-709-209-168	Sequence 168, App
c	40	13	4.0	39	1	US-08-303-275-56	Sequence 56, Appl
c	41	13	4.0	39	1	US-08-458-101-168	Sequence 168, App
	42	13	4.0	39	2	US-09-028-361A-19	Sequence 19, Appl
c	43	13	4.0	40	1	US-08-199-507B-39	Sequence 39, Appl
c	44	13	4.0	40	1	US-08-441-828-39	Sequence 39, Appl
	45	13	4.0	41	3	US-08-930-503A-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT   1
US-08-816-977-12
; Sequence 12, Application US/08816977
; Patent No. 6080400
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, James A.
;    APPLICANT:  Byrne, Lisa M.
;    APPLICANT:  Pugh, Charles S.G.
;    TITLE OF INVENTION:  Prevention And Treatment Of
;    TITLE OF INVENTION:  Verotoxin-Induced Disease
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,977
;      FILING DATE:  13-MAR-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MacKnight, Kamrin T.
;      REGISTRATION NUMBER:  38,230
;      REFERENCE/DOCKET NUMBER:  OPHD-02450
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-816-977-12


  Query Match             4.9%;  Score 16;  DB 3;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 aaaataattattttta 204
           ||||||||||||||||
```

```
Db      9 AAAATAATTATTTTTA 24


RESULT   2
US-08-629-600-16
; Sequence 16, Application US/08629600
; Patent No. 5783196
;  GENERAL INFORMATION:
;    APPLICANT:  NORIEGA, Fernando
;    APPLICANT:  LEVINE, Myron M.
;    TITLE OF INVENTION:  GUA MUTANTS OF SHIGELLA
;    TITLE OF INVENTION:  AND VACCINES CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W., Suite 800
;      CITY:  Washington, D.C.
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,600
;      FILING DATE:  9-APR-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KIT, Gordon
;      REGISTRATION NUMBER:  30,764
;      REFERENCE/DOCKET NUMBER:  A-6765
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
US-08-629-600-16


  Query Match             4.9%;  Score 16;  DB 1;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattattttta 204
          ||||||||||||||||
Db     14 AAAATAATTATTTTTA 29


RESULT   3
US-08-120-827-66
; Sequence 66, Application US/08120827
; Patent No. 5525495
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/120,827
;      FILING DATE:  15-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5525495man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  RNA (genomic)
US-08-120-827-66


  Query Match             4.6%;  Score 15;  DB 1;  Length 27;
  Best Local Similarity   33.3%;  Pred. No. 7.9e+02;
  Matches    5;  Conservative   10;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    197 tatttttatttttaaa 211
          :|:::::|::::|||
Db     12 UAUUUUUAUUUUAAA 26


RESULT   4
US-08-478-675-66
; Sequence 66, Application US/08478675
; Patent No. 5773246
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,675
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/120,827
;      FILING DATE:  15-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5773246man F.
```

```
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  RNA (genomic)
US-08-478-675-66


  Query Match             4.6%;  Score 15;  DB 1;  Length 27;
  Best Local Similarity   33.3%;  Pred. No. 7.9e+02;
  Matches    5;  Conservative   10;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    197 atttttattttaaa 211
          :|:::::|::::|||
Db     12 UAUUUUUAUUUUAAA 26


RESULT   5
PCT-US92-10792-3
; Sequence 3, Application PC/TUS9210792
;  GENERAL INFORMATION:
;    APPLICANT:  Jayasena, Sumedha D.
;    APPLICANT:  Johnston, Brian H.
;    TITLE OF INVENTION:  Triple Helix Formation at
;    TITLE OF INVENTION:  (PuNPyN)-(PuNPyN) Tracts
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SRI International
;      STREET:  333 Ravenswood Avenue
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/10792
;      FILING DATE:  19921211
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/826,934
;      FILING DATE:  21-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/808,452
;      FILING DATE:  13-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fabian, Gary R.
;      REGISTRATION NUMBER:  33,875
;      REFERENCE/DOCKET NUMBER:  P-3141
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 859-4550
;      TELEFAX:  (415) 859-3880
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  OLIGONUCLEOTIDE III, FIGURE 8
PCT-US92-10792-3


  Query Match             4.6%;  Score 15;  DB 5;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    212 tttgttaaaaatttg 226
          |||||||||||||||
Db      5 TTTGTTAAAAATTTG 19


RESULT   6
US-08-256-261-29/c
; Sequence 29, Application US/08256261
; Patent No. 5801037
;  GENERAL INFORMATION:
;    APPLICANT:  Behnke, Detlef
;    APPLICANT:  Schlott, Bernhard
;    APPLICANT:  Albrecht, Sybille
;    APPLICANT:  G hrs, Karl-Heinz
;    APPLICANT:  Hartmann, Manfred
;    TITLE OF INVENTION:  Expression of signal-peptide-free
;    TITLE OF INVENTION:  staphylokinases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/256,261
;      FILING DATE:
;      CLASSIFICATION:  435
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "synthetic DNA from
;      DESCRIPTION:  oligonucleotide synthesis"
US-08-256-261-29


  Query Match             4.6%;  Score 15;  DB 1;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 7.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    154 aaaggaaaatataaa 168
          |||||||||||||||
Db     30 AAAGGAAAATATAAA 16


RESULT   7
US-08-852-299-29/c
; Sequence 29, Application US/08852299
; Patent No. 6010897
;  GENERAL INFORMATION:
;    APPLICANT:  Behnke, Detlef
;    APPLICANT:  Schlott, Bernhard
;    APPLICANT:  Albrecht, Sybille
;    APPLICANT:  G hrs, Karl-Heinz
;    APPLICANT:  Hartmann, Manfred
;    TITLE OF INVENTION:  Expression of signal-peptide-free
```

```
;    TITLE OF INVENTION:  staphylokinases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,299
;      FILING DATE:  17-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/256,261
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "synthetic DNA from
;      DESCRIPTION:  oligonucleotide synthesis"
US-08-852-299-29


  Query Match             4.6%;  Score 15;  DB 3;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 7.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    154 aaaggaaaatataaa 168
          |||||||||||||||
Db     30 AAAGGAAAATATAAA 16


RESULT   8
US-09-205-204-20/c
; Sequence 20, Application US/09205204
; Patent No. 5958772
; GENERAL INFORMATION:
;  APPLICANT: C. Frank Bennett
;  APPLICANT: Elizabeth J. Ackermann
;  APPLICANT: Lex M. Cowsert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF CELLULAR INHIBITOR OF APOPTOSIS-1 EXPRESS
;  FILE REFERENCE: RTS-0020
;  CURRENT APPLICATION NUMBER: US/09/205,204
;  CURRENT FILING DATE: 1998-12-03
;  NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 20
;    LENGTH: 18
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Antisense Oligonucleotide
US-09-205-204-20


  Query Match             4.3%;  Score 14;  DB 2;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    160 aaatataaataaaa 173
          ||||||||||||||
Db     18 AAATATAAATAAAA 5


RESULT   9
US-08-816-977-12/c
; Sequence 12, Application US/08816977
; Patent No. 6080400
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, James A.
;    APPLICANT:  Byrne, Lisa M.
;    APPLICANT:  Pugh, Charles S.G.
;    TITLE OF INVENTION:  Prevention And Treatment Of
;    TITLE OF INVENTION:  Verotoxin-Induced Disease
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,977
;      FILING DATE:  13-MAR-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MacKnight, Kamrin T.
;      REGISTRATION NUMBER:  38,230
;      REFERENCE/DOCKET NUMBER:  OPHD-02450
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-816-977-12


  Query Match             4.3%;  Score 14;  DB 3;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     22 AAAATAATTATTTT 9


RESULT  10
US-08-629-001A-31
; Sequence 31, Application US/08629001A
; Patent No. 5858661
;  GENERAL INFORMATION:
;    APPLICANT:  Shiloh, Yosef
;    TITLE OF INVENTION:  ATAXIA-TELANGIECTASIA GENE AND ITS
;    TITLE OF INVENTION:  GENOMIC ORGANIZATION
;    NUMBER OF SEQUENCES:  139
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 5858661thwestern Hwy.
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,001A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohn, Kenneth I.
;      REGISTRATION NUMBER:  30,955
;      REFERENCE/DOCKET NUMBER:  2290.00032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 539-5050
;      TELEFAX:  (810) 539-5055
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-629-001A-31


  Query Match             4.3%;  Score 14;  DB 2;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttta 209
          ||||||||||||||
Db     14 TTATTTTTATTTTA 27


RESULT  11
US-08-642-274D-110
; Sequence 110, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
;  APPLICANT:  Shiloh, Yosef
;  TITLE OF INVENTION:  MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
;  TITLE OF INVENTION:  SCREEN FOR A PARTIAL A-T PHENOTYPE
;  FILE REFERENCE:  229000033
;  CURRENT APPLICATION NUMBER: US/08/642,274D
;  CURRENT FILING DATE:  1996-05-03
;  NUMBER OF SEQ ID NOS:  220
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 110
;   LENGTH: 30
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:intronic
;   OTHER INFORMATION: sequence
US-08-642-274D-110


  Query Match             4.3%;  Score 14;  DB 4;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttta 209
          ||||||||||||||
Db     14 ttatttttatttta 27


RESULT  12
US-08-629-600-16/c
; Sequence 16, Application US/08629600
; Patent No. 5783196
;  GENERAL INFORMATION:
;    APPLICANT:  NORIEGA, Fernando
;    APPLICANT:  LEVINE, Myron M.
;    TITLE OF INVENTION:  GUA MUTANTS OF SHIGELLA
;    TITLE OF INVENTION:  AND VACCINES CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W., Suite 800
;      CITY:  Washington, D.C.
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,600
;      FILING DATE:  9-APR-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KIT, Gordon
;      REGISTRATION NUMBER:  30,764
;      REFERENCE/DOCKET NUMBER:  A-6765
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
US-08-629-600-16


  Query Match             4.3%;  Score 14;  DB 1;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     27 AAAATAATTATTTT 14


RESULT  13
US-08-403-853-8
; Sequence 8, Application US/08403853
; Patent No. 5844094
;  GENERAL INFORMATION:
;    APPLICANT:  HUDSON, Peter J.
;    APPLICANT:  LAH, Maria
;    APPLICANT:  KORRT, Alex A.
;    APPLICANT:  IRVING, Robert A.
;    APPLICANT:  ATWELL, John L.
;    APPLICANT:  MALBY, Robyn L.
;    APPLICANT:  POWER, Barbara E.
;    APPLICANT:  COLMAN, Peter M.
;    TITLE OF INVENTION:  TARGET BINDING POLYPEPTIDE
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,853
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/AU93/00491
;      FILING DATE:  24-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PL 4973
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  16786/189/CHAC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-403-853-8


  Query Match             4.3%;  Score 14;  DB 2;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 aaatttataattat 96
           ||||||||||||||
Db       1 AAATTTATAATTAT 14


RESULT  14
US-08-847-844A-113
; Sequence 113, Application US/08847844A
; Patent No. 6150160
;  GENERAL INFORMATION:
;    APPLICANT:  KAZAZIAN JR., HAIG H.
;    APPLICANT:  BOEKE, JEF D.
;    APPLICANT:  MORAN, JOHN V.
;    APPLICANT:  DOMBROSKI, BETH A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS OF USE OF
;    TITLE OF INVENTION:  MAMMALIAN RETROTRANSPOSONS
;    NUMBER OF SEQUENCES:  137
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND FL.
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/847,844A
;      FILING DATE:  28-APR-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/749,805
;      FILING DATE:  16-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/006,831
;      FILING DATE:  16-NOV-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOYLE LEARY Ph.D., KATHRYN
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9596-23U2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-567-2020
;      TELEFAX:  215-567-2991
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-847-844A-113


  Query Match             4.0%;  Score 13;  DB 3;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     150 ttaaaaaggaaaa 162
           |||||||||||||
Db       5 TTAAAAAGGAAAA 17


RESULT  15
US-08-686-968C-13
; Sequence 13, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
;  APPLICANT: Cochran, Mark D.
;  APPLICANT:  Junker, David E.
;  TITLE OF INVENTION: Recombinant Swinepox Virus
;  FILE REFERENCE: 39119-H/JML
;  CURRENT APPLICATION NUMBER: US/08/686,968C
;  CURRENT FILING DATE:  1996-07-25
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 18
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: primer
US-08-686-968C-13


  Query Match             4.0%;  Score 13;  DB 4;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     231 aattgaattgtaa 243
           |||||||||||||
Db       5 aattgaattgtaa 17


RESULT  16
US-09-288-461-79
; Sequence 79, Application US/09288461
; Patent No. 6159694
; GENERAL INFORMATION:
;  APPLICANT: Karras, James G.
;  TITLE OF INVENTION: Antisense Oligonucleotide Modulation of STAT3
;  TITLE OF INVENTION: Expression
;  FILE REFERENCE: ISPH-0338
;  CURRENT APPLICATION NUMBER: US/09/288,461
;  CURRENT FILING DATE: 1999-04-08
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Sequence
US-09-288-461-79


  Query Match             4.0%;  Score 13;  DB 3;  Length 20;
```

```
  Best Local Similarity   100.0%;  Pred. No. 5.2e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    173 atttagataaaag 185
          |||||||||||||
Db      2 atttagataaaag 14


RESULT  17
US-08-691-045-61/c
; Sequence 61, Application US/08691045
; Patent No. 6015664
;  GENERAL INFORMATION:
;    APPLICANT:  Henrickson, Kelly J.
;    APPLICANT:  Fan, Jiang (n.m.i.)
;    TITLE OF INVENTION:  VIRUS ASSAY METHOD
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,045
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  650053.91037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414) 271-3552
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Oligonucleotide
US-08-691-045-61


  Query Match             4.0%;  Score 13;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     94 tatatgataagta 106
          |||||||||||||
Db     17 TATATGATAAGTA 5


RESULT  18
US-08-672-215-1/c
; Sequence 1, Application US/08672215
; Patent No. 6020121
;  GENERAL INFORMATION:
;    APPLICANT:  Ying Bao, Amy Boggs, Pamela R. Contag,
;    APPLICANT:  Nancy A. Federspiel, Alan Herbert,
;    APPLICANT:  Scott J. Hecker, Francois Malouin
;    TITLE OF INVENTION:  INHIBITORS OF REGULATORY PATHWAYS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/672,215
;      FILING DATE:  June 25, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/004,626
;      FILING DATE:  September 29, 1995
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  219/158
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-672-215-1


  Query Match             4.0%;  Score 13;  DB 3;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    294 tgtaatttttatg 306
          |||||||||||||
Db     20 TGTAATTTTTATG 8


RESULT  19
US-08-120-827-64
; Sequence 64, Application US/08120827
; Patent No. 5525495
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/120,827
;      FILING DATE:  15-SEP-1993
```

```
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Oblon, No. 5525495man F.
;       REGISTRATION NUMBER:  24,618
;       REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)413-3000
;       TELEFAX:  (703)413-2220
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  64:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  RNA (genomic)
US-08-120-827-64


  Query Match             4.0%;  Score 13;  DB 1;  Length 28;
  Best Local Similarity   15.4%;  Pred. No. 5e+03;
  Matches    2;  Conservative   11;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     196 ttatttttatttt 208
           ::|:::::|::::
Db       3 UUAUUUUUAUUUU 15


RESULT  20
US-08-478-675-64
; Sequence 64, Application US/08478675
; Patent No. 5773246
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,675
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/120,827
;      FILING DATE:  15-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5773246man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  RNA (genomic)
US-08-478-675-64


  Query Match             4.0%;  Score 13;  DB 1;  Length 28;
  Best Local Similarity   15.4%;  Pred. No. 5e+03;
  Matches    2;  Conservative   11;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     196 ttatttttatttt 208
           ::|:::::|::::
Db       3 UUAUUUUUAUUUU 15


RESULT  21
US-08-629-001A-79
; Sequence 79, Application US/08629001A
; Patent No. 5858661
;  GENERAL INFORMATION:
;    APPLICANT:  Shiloh, Yosef
;    TITLE OF INVENTION:  ATAXIA-TELANGIECTASIA GENE AND ITS
;    TITLE OF INVENTION:  GENOMIC ORGANIZATION
;    NUMBER OF SEQUENCES:  139
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 5858661thwestern Hwy.
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,001A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohn, Kenneth I.
;      REGISTRATION NUMBER:  30,955
;      REFERENCE/DOCKET NUMBER:  2290.00032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 539-5050
;      TELEFAX:  (810) 539-5055
;  INFORMATION FOR SEQ ID NO:  79:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-629-001A-79


  Query Match             4.0%;  Score 13;  DB 2;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 4.9e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     272 aaaaaattatttc 284
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT  22
US-08-642-274D-158
; Sequence 158, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
;  APPLICANT:  Shiloh, Yosef
;  TITLE OF INVENTION:  MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
;  TITLE OF INVENTION:  SCREEN FOR A PARTIAL A-T PHENOTYPE
;  FILE REFERENCE:  229000033
```

```
;  CURRENT APPLICATION NUMBER: US/08/642,274D
;  CURRENT FILING DATE:  1996-05-03
;  NUMBER OF SEQ ID NOS:  220
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 158
;   LENGTH: 30
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:intronic
;   OTHER INFORMATION: sequence
US-08-642-274D-158


  Query Match             4.0%;  Score 13;  DB 4;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 4.9e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    272 aaaaaattatttc 284
          |||||||||||||
Db     15 aaaaaattatttc 27


RESULT  23
US-08-330-638D-5/c
; Sequence 5, Application US/08330638D
; Patent No. 5731425
;  GENERAL INFORMATION:
;    APPLICANT:  Brizzard, Billy
;    APPLICANT:  Bianca, Darlene
;    APPLICANT:  Chubet, Richard
;    APPLICANT:  Vizard, Douglas
;    APPLICANT:  Hopp, Thomas
;    TITLE OF INVENTION:  POLYPEPTIDE SURFACE
;    TITLE OF INVENTION:  MARKER FOR CELLS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eastman Kodak Company,
;      ADDRESSEE:  Patent Legal Staff
;      STREET:  343 State Street
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14650-2201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch,
;      MEDIUM TYPE:  1.44 MB storage, (Hewlett Packard)
;      COMPUTER:  HP Vectra
;      OPERATING SYSTEM:  MS-DOS Version 6.0
;      SOFTWARE:  WORD FOR WINDOWS
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/330,638D
;      FILING DATE:  28 OCT 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:  NONE
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kiernan, Anne B.
;      REGISTRATION NUMBER:  36,566
;      REFERENCE/DOCKET NUMBER:  71255
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 588-2405
;      TELEFAX:  (716) 477-4646
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 BASES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  SYNTHETIC OLIGONUCLEOTIDE
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:  SYNTHETICALLY PREPARED
;    IMMEDIATE SOURCE:  SYNTHETICALLY PREPARED
;    PUBLICATION INFORMATION:  NONE
US-08-330-638D-5


  Query Match             4.0%;  Score 13;  DB 1;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 4.9e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    124 gtgttctcggggg 136
          |||||||||||||
Db     28 GTGTTCTCGGGGG 16


RESULT  24
US-08-906-746A-5/c
; Sequence 5, Application US/08906746A
; Patent No. 5945292
;  GENERAL INFORMATION:
;    APPLICANT:  Brizzard, Billy L.
;    APPLICANT:  Bianca, Darlene W.
;    APPLICANT:  Chubert, Richard G.
;    APPLICANT:  Vizard, Douglas L.
;    APPLICANT:  Hopp, Thomas P.
;    TITLE OF INVENTION:  Method of Identifying Cells with
;    TITLE OF INVENTION:  Polypeptide Surface Marker
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Senniger, Powers, Leavitt & Roedel
;      STREET:  One Metropolitan Square- 16th Floor
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/906,746A
;      FILING DATE:  06-AUG-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stone, Paul A.
;      REGISTRATION NUMBER:  38,628
;      REFERENCE/DOCKET NUMBER:  SGM 6874
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-231-5400
;      TELEFAX:  314-231-4342
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-906-746A-5


  Query Match             4.0%;  Score 13;  DB 2;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 4.9e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    124 gtgttctcggggg 136
          |||||||||||||
Db     28 GTGTTCTCGGGGG 16


RESULT  25
US-08-247-809A-14/c
; Sequence 14, Application US/08247809A
```

```
; Patent No. 5569823
;  GENERAL INFORMATION:
;    APPLICANT:  Peter H. Schreier; Klaus Stenzel; Gunter Adam;
;    APPLICANT:  Edgar Maiss
;    TITLE OF INVENTION:  DEOXYRIBONUCLEIC ACIDS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 2.0 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  NEC Powermate 1 Plus
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/247,809A
;      FILING DATE:  May 23, 1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  P 43 178 45.6 (Germany)
;      FILING DATE:  May 28, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 9049-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-247-809A-14


  Query Match             4.0%;  Score 13;  DB 1;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    156 aggaaaatataaa 168
          |||||||||||||
Db     24 AGGAAAATATAAA 12


RESULT  26
US-08-711-728-14/c
; Sequence 14, Application US/08711728
; Patent No. 5973135
;  GENERAL INFORMATION:
;    APPLICANT:  Peter H. Schreier; Klaus Stenzel; Gunter Adam;
;    APPLICANT:  Edgar Maiss
;    TITLE OF INVENTION:  DEOXYRIBONUCLEIC ACIDS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 2.0 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  NEC Powermate 1 Plus
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/711,728
;      FILING DATE:  03-SEPT-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/247,809
;      FILING DATE:  23-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 43178456
;      FILING DATE:  28-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 9049.1-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-711-728-14


  Query Match             4.0%;  Score 13;  DB 2;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    156 aggaaaatataaa 168
          |||||||||||||
Db     24 AGGAAAATATAAA 12


RESULT  27
US-08-097-554A-45
; Sequence 45, Application US/08097554A
; Patent No. 5869312
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran Ph.D., Mark D
;    APPLICANT:  Junker M.S., David E
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  112
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/097,554A
;      FILING DATE:  July 22, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)977-9550
;      TELEFAX:  (212)664-0525
;      TELEX:  422523
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Swinepox virus
;      STRAIN:  Kasza
;      INDIVIDUAL ISOLATE:  S-SPV-001
;    IMMEDIATE SOURCE:
;      CLONE:  515-85.1
;    POSITION IN GENOME:
;      MAP POSITION:  ~23.2
;      UNITS:  %G
US-08-097-554A-45


  Query Match             4.0%;  Score 13;  DB 2;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  28
US-08-484-575A-10
; Sequence 10, Application US/08484575A
; Patent No. 5925358
;  GENERAL INFORMATION:
;    APPLICANT:  Mark D. Cochran and David E. Junker
;    TITLE OF INVENTION:  Recombinant Fowlpox Viruses and Uses Thereof
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,575A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White Esq, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)278-0450
;      TELEFAX:  (212)391-0525
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-484-575A-10


  Query Match             4.0%;  Score 13;  DB 2;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  29
US-08-477-459-10
; Sequence 10, Application US/08477459
; Patent No. 6001369
;  GENERAL INFORMATION:
;    APPLICANT:  Mark D. Cochran
;    TITLE OF INVENTION:  Recombinant Fowlpox Viruses and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,459
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White Esq, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-477-459-10


  Query Match             4.0%;  Score 13;  DB 3;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  30
US-08-480-640A-45
; Sequence 45, Application US/08480640A
; Patent No. 6033904
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  Junker, David E.
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  225
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,640A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P
;       REGISTRATION NUMBER:  28,678
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278-0400
;       TELEFAX:  (212) 391-0525
;   INFORMATION FOR SEQ ID NO:  45:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  37 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Swinepox virus
;       STRAIN:  Kasza
;       INDIVIDUAL ISOLATE:  S-SPV-001
;     IMMEDIATE SOURCE:
;       CLONE:  515-85.1
;     POSITION IN GENOME:
;       MAP POSITION:  ~23.2
;       UNITS:  %G
US-08-480-640A-45


  Query Match             4.0%;  Score 13;  DB 3;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  31
US-08-479-869-10
; Sequence 10, Application US/08479869
; Patent No. 6123949
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran Ph.D, Mark D
;    TITLE OF INVENTION:  Recombinant Fowlpox Virus S-FPV-043 and
;    TITLE OF INVENTION:  Uses Thereof
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,869
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/024,156
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White Esq, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)977-9550
;      TELEFAX:  (212)664-0525
;      TELEX:  422523
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-479-869-10


  Query Match             4.0%;  Score 13;  DB 3;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  32
US-08-295-802-45
; Sequence 45, Application US/08295802
; Patent No. 6127163
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran Ph.D., Mark D
;    APPLICANT:  Junker M.S., David E
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  188
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/295,802
;      FILING DATE:  Herewith
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)977-9550
;      TELEFAX:  (212)664-0525
;      TELEX:  422523
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Swinepox virus
;      STRAIN:  Kasza
;      INDIVIDUAL ISOLATE:  S-SPV-001
;    IMMEDIATE SOURCE:
;      CLONE:  515-85.1
;    POSITION IN GENOME:
;      MAP POSITION:  ~23.2
```

```
;      UNITS:  %G
US-08-295-802-45


  Query Match             4.0%;  Score 13;  DB 3;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  33
US-08-486-414-10
; Sequence 10, Application US/08486414B
; Patent No. 6136318
; GENERAL INFORMATION:
;  APPLICANT: Cochran, Mark D.
;  APPLICANT: Junker, David E.
;  TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUSES AND USES THEREOF
;  FILE REFERENCE: 42771D
;  CURRENT APPLICATION NUMBER: US/08/486,414B
;  CURRENT FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 37
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Linker
US-08-486-414-10


  Query Match             4.0%;  Score 13;  DB 3;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 aattgaattgtaa 22


RESULT  34
US-08-488-237A-45
; Sequence 45, Application US/08488237A
; Patent No. 6251403
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  Junker, David E.
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  225
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,237A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Swinepox virus
;      STRAIN:  Kasza
;      INDIVIDUAL ISOLATE:  S-SPV-001
;    IMMEDIATE SOURCE:
;      CLONE:  515-85.1
;    POSITION IN GENOME:
;      MAP POSITION:  ~23.2
;      UNITS:  %G
US-08-488-237A-45


  Query Match             4.0%;  Score 13;  DB 4;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  35
PCT-US94-01826A-10
; Sequence 10, Application PC/TUS9401826A
;  GENERAL INFORMATION:
;    APPLICANT:  Syntro Corporation, et al.
;    TITLE OF INVENTION:  Recombinant Fowlpox Virus S-FPV-043 and Uses  Thereof
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/01826A
;      FILING DATE:  28-FEB-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White Esq, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)977-9550
;      TELEFAX:  (212)664-0525
;      TELEX:  422523
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US94-01826A-10
```

```
  Query Match             4.0%;  Score 13;  DB 5;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  36
PCT-US94-02252A-10
; Sequence 10, Application PC/TUS9402252A
;  GENERAL INFORMATION:
;    APPLICANT:  Syntro Corporation, et al.
;    TITLE OF INVENTION:  Recombinant Fowlpox Viruses and Uses Thereof
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/02252A
;      FILING DATE:  28-FEB-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White Esq, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)977-9550
;      TELEFAX:  (212)664-0525
;      TELEX:  422523
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US94-02252A-10


  Query Match             4.0%;  Score 13;  DB 5;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    231 aattgaattgtaa 243
          |||||||||||||
Db     10 AATTGAATTGTAA 22


RESULT  37
PCT-US96-00547-40/c
; Sequence 40, Application PC/TUS9600547
;  GENERAL INFORMATION:
;    APPLICANT:  Virogenetics Corporation
;    TITLE OF INVENTION:  RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00547
;      FILING DATE:  12-JAN-1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/372,664
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  38 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-00547-40


  Query Match             4.0%;  Score 13;  DB 5;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    293 ctgtaatttttat 305
          |||||||||||||
Db     25 CTGTAATTTTTAT 13


RESULT  38
US-08-105-483-168/c
; Sequence 168, Application US/08105483
; Patent No. 5494807
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  462
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/105,483
;      FILING DATE:  12-AUG-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/847,951
;      FILING DATE:  06-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2400
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;   INFORMATION FOR SEQ ID NO:  168:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  39 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-105-483-168


  Query Match             4.0%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     293 ctgtaatttttat 305
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT  39
US-08-709-209-168/c
; Sequence 168, Application US/08709209
; Patent No. 5762938
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  462
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/709,209
;      FILING DATE:  21-AUG-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/105,483
;      FILING DATE:  12-AUG-1993
;      APPLICATION NUMBER:  US 07/847,951
;      FILING DATE:  06-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  168:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-709-209-168


  Query Match             4.0%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     293 ctgtaatttttat 305
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT  40
US-08-303-275-56/c
; Sequence 56, Application US/08303275
; Patent No. 5766598
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  IMMUNODEFICIENCY VIRUS RECOMBINANT
;    TITLE OF INVENTION:  POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  205
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/303,275
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/897,382
;      FILING DATE:  11-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-303-275-56


  Query Match             4.0%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     293 ctgtaatttttat 305
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT  41
US-08-458-101-168/c
; Sequence 168, Application US/08458101
; Patent No. 5766599
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Perkus, Marion E.
;    APPLICANT:  Taylor, Jill
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  No. 5766599ton, Elizabeth K.
;    APPLICANT:  Riviere, Michel
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Limbach, Keith J.
;    APPLICANT:  Johnson, Gerard P.
```

```
;    APPLICANT:  Pincus, Steven E.
;    APPLICANT:  Cox, William I.
;    APPLICANT:  Audonnet, Jean-Christophe Francis
;    APPLICANT:  Gettig, Russell Robert
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  467
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,101
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2740
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  168:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-458-101-168


  Query Match             4.0%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     293 ctgtaatttttat 305
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT  42
US-09-028-361A-19
; Sequence 19, Application US/09028361A
; Patent No. 5962296
;  GENERAL INFORMATION:
;    APPLICANT:  ETTWILLER, LAURENCE
;    APPLICANT:  XU, SHUANG-YONG
;    TITLE OF INVENTION:  METHOD FOR CLONING AND
;    TITLE OF INVENTION:  PRODUCING THERMOMICROBIUM ROSEUM DNA
;    TITLE OF INVENTION:  POLYMERASE I IN E. COLI
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  New England Biolabs, Inc.
;      STREET:  32 Tozer Road
;      CITY:  Beverly
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  01915
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/028,361A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Gregory D.
;      REGISTRATION NUMBER:  30901
;      REFERENCE/DOCKET NUMBER:  NEB-134
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  978-927-5054
;      TELEFAX:  978-927-1705
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-028-361A-19


  Query Match             4.0%;  Score 13;  DB 2;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      62 atatattaattat 74
           |||||||||||||
Db      12 ATATATTAATTAT 24


RESULT  43
US-08-199-507B-39/c
; Sequence 39, Application US/08199507B
; Patent No. 5472841
;  GENERAL INFORMATION:
;    APPLICANT:  JAYASENA, S. AND GOLD, L.
;    TITLE OF INVENTION:  NUCLEIC ACID LIGAND INHIBITORS
;    TITLE OF INVENTION:  OF HUMAN NEUTROPHIL ELASTASE
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 E. Prentice Avenue, Suite 200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 Kb storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/199,507B
;      FILING DATE:  22 FEBRUARY 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/714,131
;      FILING DATE:  10-JUNE-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/536,428
;      FILING DATE:  11-JUNE-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX13
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  40 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-199-507B-39


  Query Match             4.0%;  Score 13;  DB 1;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    157 ggaaaatataaat 169
          |||||||||||||
Db     22 GGAAAATATAAAT 10


RESULT  44
US-08-441-828-39/c
; Sequence 39, Application US/08441828
; Patent No. 5734034
;  GENERAL INFORMATION:
;    APPLICANT:  JAYASENA, S. AND GOLD, L.
;    TITLE OF INVENTION:  NUCLEIC ACID LIGAND INHIBITORS
;    TITLE OF INVENTION:  OF HUMAN NEUTROPHIL ELASTASE
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 E. Prentice Avenue, Suite 200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,828
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/199,507
;      FILING DATE:  22 FEBRUARY 1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/714,131
;      FILING DATE:  10-JUNE-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/536,428
;      FILING DATE:  11-JUNE-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX13
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-441-828-39


  Query Match             4.0%;  Score 13;  DB 1;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    157 ggaaaatataaat 169
          |||||||||||||
```

```
Db     22 GGAAAATATAAAT 10


RESULT  45
US-08-930-503A-11
; Sequence 11, Application US/08930503A
; Patent No. 6132731
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MURINE LEUKEMIA VIRUS VECTORS (AS AMENDED)
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  WENDEROTH, LIND & PONACK, L.L.P.
;      STREET:  2033 K Street, N.W. - Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/930,503A
;      FILING DATE:  October 8, 1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9506782.3
;      FILING DATE:  01-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/GB96/00776
;      FILING DATE:  01-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  41 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-930-503A-11


  Query Match             4.0%;  Score 13;  DB 3;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    251 ttcagggggaat 263
          |||||||||||||
Db      2 TTCAGGGGGAAT 14


Search completed: January 24, 2002, 03:24:24
Job time: 3711 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:19:03 ; Search time 1494.92 Seconds
(without alignments)
3608.606 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 541028

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.		Score	Query Match	Length	DB	ID	Description
1		16	4.9	29	6	AR099868	AR099868 Sequence
2		16	4.9	32	6	E27913	E27913 Method for
c 3		16	4.9	36	3	CEANONYFR	X97532 C.elegans D
4		16	4.9	36	6	AR019036	AR019036 Sequence
5		15	4.6	27	6	AR014030	AR014030 Sequence
6		15	4.6	27	6	I21980	I21980 Sequence 66
c 7		15	4.6	32	6	AR037189	AR037189 Sequence
c 8		14	4.3	18	6	AR076353	AR076353 Sequence
c 9		14	4.3	29	6	AR099868	AR099868 Sequence
10		14	4.3	30	6	AR028182	AR028182 Sequence
11		14	4.3	30	6	AR138585	AR138585 Sequence
c 12		14	4.3	32	6	E27913	E27913 Method for
c 13		14	4.3	36	6	AR019036	AR019036 Sequence
14		14	4.3	37	6	AR063204	AR063204 Sequence
c 15		14	4.3	45	6	AX049973	AX049973 Sequence
16		14	4.3	45	6	AX049974	AX049974 Sequence
c 17		14	4.3	45	6	AX099855	AX099855 Sequence
18		14	4.3	45	6	AX099856	AX099856 Sequence
c 19		14	4.3	45	6	AX137975	AX137975 Sequence
20		14	4.3	45	6	AX137976	AX137976 Sequence
c 21		14	4.3	50	6	AX159856	AX159856 Sequence
22		13	4.0	18	6	AR146953	AR146953 Sequence
23		13	4.0	20	6	AR121058	AR121058 Sequence
c 24		13	4.0	20	6	AX076045	AX076045 Sequence
25		13	4.0	23	6	A97479	A97479 Sequence 35
26		13	4.0	24	6	AX093544	AX093544 Sequence
27		13	4.0	24	6	AX164353	AX164353 Sequence
c 28		13	4.0	24	12	AB069100	AB069100 Synthetic
c 29		13	4.0	25	6	AX042574	AX042574 Sequence
c 30		13	4.0	25	6	AX043268	AX043268 Sequence
31		13	4.0	26	6	AX039624	AX039624 Sequence
32		13	4.0	26	6	AX039654	AX039654 Sequence
33		13	4.0	28	6	AR014028	AR014028 Sequence
34		13	4.0	28	6	I21978	I21978 Sequence 64
35		13	4.0	29	6	AX012366	AX012366 Sequence
36		13	4.0	29	6	E59972	E59972 Highly acti
37		13	4.0	30	6	AR028230	AR028230 Sequence
38		13	4.0	30	6	AR138633	AR138633 Sequence
c 39		13	4.0	30	6	AX063379	AX063379 Sequence
c 40		13	4.0	31	6	I95122	I95122 Sequence 5
41		13	4.0	33	5	XELARSE59	K01606 Xenopus lae
c 42		13	4.0	36	6	A41027	A41027 Sequence 14
c 43		13	4.0	36	6	AR082586	AR082586 Sequence
44		13	4.0	36	6	AX167671	AX167671 Sequence
c 45		13	4.0	36	6	I28261	I28261 Sequence 14

ALIGNMENTS

RESULT 1
AR099868
LOCUS AR099868 29 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 12 from patent US 6080400.
ACCESSION AR099868
VERSION AR099868.1 GI:12810316
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 29)
AUTHORS Williams,J.A. and Byrne,L.Marie.
TITLE Compositions for the prevention and treatment of verotoxin-induced disease
JOURNAL Patent: US 6080400-A 12 27-JUN-2000;
FEATURES Location/Qualifiers
source 1. .29
/organism="unknown"
BASE COUNT 11 a 2 c 5 g 11 t
ORIGIN

```
  Query Match             4.9%;  Score 16;  DB 6;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 9.2e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattattttta 204
          ||||||||||||||||
Db      9 AAAATAATTATTTTTA 24


RESULT   2
E27913
LOCUS       E27913         32 bp    DNA             PAT       07-FEB-2001
DEFINITION  Method for detecting foreign DNA fragment insert in Vero toxin
            gene.
ACCESSION   E27913
VERSION     E27913.1  GI:13020766
KEYWORDS    JP 1999243996-A/3.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Masahiro,K.Y.N.N. and Kawamura,K.S.
  TITLE     Method for detecting foreign DNA fragment insert in Vero toxin gene
  JOURNAL   Patent: JP 1999243996-A 3 14-SEP-1999;
            TOYOBO CO LTD
COMMENT     OS   Unidentified
            PN   JP 1999243996-A/3
            PD   14-SEP-1999
            PF   27-FEB-1998 JP 1998047677
            PR
            PI   MASAHIRO KUSUMOTO,YOSHIAKI NISHIYA,YOSHIHISA KAWAMURA, PI
            KUNIHIRO SHINAGAWA
            PC   C12Q1/68,C12N15/09//(C12N15/09,C12R1:185),C12N15/00,
            (C12N15/00, PC   C12R1:185)
            CC   Strandedness: Both;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .32
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       12 a      2 c      4 g     14 t
ORIGIN


  Query Match             4.9%;  Score 16;  DB 6;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 9.2e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattattttta 204
          ||||||||||||||||
Db      4 AAAATAATTATTTTTA 19


RESULT   3
CEANONYFR/c
LOCUS       CEANONYFR      36 bp    DNA             INV       06-MAY-1997
DEFINITION  C.elegans DNA fragment with rearrangement site.
ACCESSION   X97532
VERSION     X97532.1  GI:1296496
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Wicky,C., Villeneuve,A.M., Lauper,N., Codourey,L., Tobler,H. and
            Muller,F.
  TITLE     Telomeric repeats (TTAGGC)n are sufficient for chromosome capping
            function in Caenorhabditis elegans
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (17), 8983-8988 (1996)
  MEDLINE   96392352
REFERENCE   2  (bases 1 to 36)
  AUTHORS   Wicky,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-1996) C. Wicky, University of British Columbia,
            Medical Genetics, 6174 University Boulevard, Vancouver, B.C. V6T
            1Z3, CANADA
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="BP1"
     misc_feature    1. .36
                     /note="sequence with rearrangement site"
BASE COUNT        9 a      5 c      4 g     18 t
ORIGIN


  Query Match             4.9%;  Score 16;  DB 3;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 9.1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    169 taaaatttagataaaa 184
          ||||||||||||||||
Db     23 TAAAATTTAGATAAAA 8


RESULT   4
AR019036
LOCUS       AR019036       36 bp    DNA             PAT       05-DEC-1998
DEFINITION  Sequence 16 from patent US 5783196.
ACCESSION   AR019036
VERSION     AR019036.1  GI:3974150
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Noriega,F.R. and Levine,M.M.
  TITLE     Gua mutants of shigella spp. and vaccines containing the same
  JOURNAL   Patent: US 5783196-A 16 21-JUL-1998;
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="unknown"
BASE COUNT       11 a      3 c     10 g     12 t
ORIGIN


  Query Match             4.9%;  Score 16;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 9.1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattattttta 204
          ||||||||||||||||
Db     14 AAAATAATTATTTTTA 29


RESULT   5
AR014030
LOCUS       AR014030       27 bp    DNA             PAT       05-DEC-1998
DEFINITION  Sequence 66 from patent US 5773246.
ACCESSION   AR014030
VERSION     AR014030.1  GI:3971484
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Keene,J.D., Levine,T. and Gao,F.
  TITLE     Methods and compositions useful in the recognition, binding and
            expression of ribonucleic acids involved in cell growth, neoplasia
```

```
          and immunoregulation
  JOURNAL   Patent: US 5773246-A 66 30-JUN-1998;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
BASE COUNT        7 a      2 c      2 g     16 t
ORIGIN
```

```
  Query Match             4.6%;  Score 15;  DB 6;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+04;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    197 atttttattttaaa 211
          |||||||||||||||
Db     12 TATTTTTATTTTAAA 26
```

```
RESULT   6
I21980
LOCUS       I21980         27 bp    DNA             PAT       07-OCT-1996
DEFINITION  Sequence 66 from patent US 5525495.
ACCESSION   I21980
VERSION     I21980.1  GI:1602334
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Keene,J.D., Levine,T. and Gao,F.
  TITLE     Methods and compositions useful in the recognition, binding and
            expression of ribonucleic acids involved in cell growth, neoplasia
            and immunoregulation
  JOURNAL   Patent: US 5525495-A 66 11-JUN-1996;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
BASE COUNT        7 a      2 c      2 g     16 t
ORIGIN
```

```
  Query Match             4.6%;  Score 15;  DB 6;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+04;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    197 atttttattttaaa 211
          |||||||||||||||
Db     12 TATTTTTATTTTAAA 26
```

```
RESULT   7
AR037189/c
LOCUS       AR037189       32 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 29 from patent US 5801037.
ACCESSION   AR037189
VERSION     AR037189.1  GI:5955045
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Behnke,D., Schlott,B., Albrecht,S., Guhrs,K. and Hartmann,M.
  TITLE     Expression of signal-peptide-free staphylokinases
  JOURNAL   Patent: US 5801037-A 29 01-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="unknown"
BASE COUNT        3 a      8 c      4 g     17 t
ORIGIN
```

```
  Query Match             4.6%;  Score 15;  DB 6;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+04;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    154 aaaggaaaatataaa 168
          |||||||||||||||
Db     30 AAAGGAAAATATAAA 16
```

```
RESULT   8
AR076353/c
LOCUS       AR076353       18 bp    DNA             PAT       30-AUG-2000
DEFINITION  Sequence 20 from patent US 5958772.
ACCESSION   AR076353
VERSION     AR076353.1  GI:10003099
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Bennett,C.Frank, Ackermann,E.J. and Cowsert,L.M.
  TITLE     Antisense inhibition of cellular inhibitor of apoptosis-1
            expression
  JOURNAL   Patent: US 5958772-A 20 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unknown"
BASE COUNT        3 a      2 c      1 g     12 t
ORIGIN
```

```
  Query Match             4.3%;  Score 14;  DB 6;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    160 aaatataaaataaaa 173
          ||||||||||||||
Db     18 AAATATAAATAAAA 5
```

```
RESULT   9
AR099868/c
LOCUS       AR099868       29 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 12 from patent US 6080400.
ACCESSION   AR099868
VERSION     AR099868.1  GI:12810316
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Williams,J.A. and Byrne,L.Marie.
  TITLE     Compositions for the prevention and treatment of verotoxin-induced
            disease
  JOURNAL   Patent: US 6080400-A 12 27-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       11 a      2 c      5 g     11 t
ORIGIN
```

```
  Query Match             4.3%;  Score 14;  DB 6;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 6.1e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     22 AAAATAATTATTTT 9
```

```
RESULT  10
AR028182
LOCUS       AR028182       30 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 31 from patent US 5858661.
```

```
ACCESSION   AR028182
VERSION     AR028182.1  GI:5940155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Ataxia-telangiectasia gene and its genomic organization
  JOURNAL   Patent: US 5858661-A 31 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT       11 a      1 c      5 g     13 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 6.1e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttta 209
          ||||||||||||||
Db     14 TTATTTTTATTTTA 27


RESULT  11
AR138585
LOCUS       AR138585       30 bp    DNA             PAT       16-JUN-2001
DEFINITION  Sequence 110 from patent US 6200749.
ACCESSION   AR138585
VERSION     AR138585.1  GI:14480930
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Mutated forms of the ataxia-telangiectasia gene and method to
            screen for a partial A-T phenotype
  JOURNAL   Patent: US 6200749-A 110 13-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT       11 a      1 c      5 g     13 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 6.1e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttta 209
          ||||||||||||||
Db     14 TTATTTTTATTTTA 27


RESULT  12
E27913/c
LOCUS       E27913         32 bp    DNA             PAT       07-FEB-2001
DEFINITION  Method for detecting foreign DNA fragment insert in Vero toxin
            gene.
ACCESSION   E27913
VERSION     E27913.1  GI:13020766
KEYWORDS    JP 1999243996-A/3.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Masahiro,K.Y.N.N. and Kawamura,K.S.
  TITLE     Method for detecting foreign DNA fragment insert in Vero toxin gene
  JOURNAL   Patent: JP 1999243996-A 3 14-SEP-1999;
            TOYOBO CO LTD
COMMENT     OS   Unidentified
            PN   JP 1999243996-A/3
            PD   14-SEP-1999
            PF   27-FEB-1998 JP 1998047677
            PR
            PI   MASAHIRO KUSUMOTO,YOSHIAKI NISHIYA,YOSHIHISA KAWAMURA, PI
            KUNIHIRO SHINAGAWA
            PC   C12Q1/68,C12N15/09//(C12N15/09,C12R1:185),C12N15/00,
            (C12N15/00, PC   C12R1:185)
            CC   Strandedness: Both;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .32
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       12 a      2 c      4 g     14 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 6.1e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     17 AAAATAATTATTTT 4


RESULT  13
AR019036/c
LOCUS       AR019036       36 bp    DNA             PAT       05-DEC-1998
DEFINITION  Sequence 16 from patent US 5783196.
ACCESSION   AR019036
VERSION     AR019036.1  GI:3974150
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Noriega,F.R. and Levine,M.M.
  TITLE     Gua mutants of shigella spp. and vaccines containing the same
  JOURNAL   Patent: US 5783196-A 16 21-JUL-1998;
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="unknown"
BASE COUNT       11 a      3 c     10 g     12 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     27 AAAATAATTATTTT 14


RESULT  14
AR063204
LOCUS       AR063204       37 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 8 from patent US 5844094.
ACCESSION   AR063204
VERSION     AR063204.1  GI:5990895
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   Hudson,P.John, Lah,M., Kortt,A.Andrew, Irving,R.Alexander,
```

```
            Atwell,J.Leslie, Malby,R.Louise, Power,B.Elaine and
            Colman,P.Malcolm.
  TITLE     Target binding polypeptide
  JOURNAL   Patent: US 5844094-A 8 01-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="unknown"
BASE COUNT       14 a      5 c      0 g     18 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 aaatttataattat 96
           ||||||||||||||
Db       1 AAATTTATAATTAT 14


RESULT  15
AX049973/c
LOCUS       AX049973       45 bp    DNA             PAT       12-JAN-2001
DEFINITION  Sequence 74 from Patent WO0070071.
ACCESSION   AX049973
VERSION     AX049973.1  GI:12226350
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Bout,A., Havenga,M.J. and Vogels,R.
  TITLE     Adenovirus derived gene delivery vehicles comprising at least one
            element of adenovirus type 35
  JOURNAL   Patent: WO 0070071-A 74 23-NOV-2000;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide TATA-plus"
BASE COUNT       14 a      5 c      7 g     19 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      21 CTGAAAATTTATAA 8


RESULT  16
AX049974
LOCUS       AX049974       45 bp    DNA             PAT       12-JAN-2001
DEFINITION  Sequence 75 from Patent WO0070071.
ACCESSION   AX049974
VERSION     AX049974.1  GI:12226351
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Bout,A., Havenga,M.J. and Vogels,R.
  TITLE     Adenovirus derived gene delivery vehicles comprising at least one
            element of adenovirus type 35
  JOURNAL   Patent: WO 0070071-A 75 23-NOV-2000;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide TATA-min"
BASE COUNT       19 a      7 c      5 g     14 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      29 CTGAAAATTTATAA 42


RESULT  17
AX099855/c
LOCUS       AX099855       45 bp    DNA             PAT       02-APR-2001
DEFINITION  Sequence 4 from Patent WO0120014.
ACCESSION   AX099855
VERSION     AX099855.1  GI:13538881
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Schouten,G.J., Vogels,R. and Opstelten,D.J.
  TITLE     Modified adenoviral vectors for use in gene therapy
  JOURNAL   Patent: WO 0120014-A 4 22-MAR-2001;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
     primer_bind     1. .45
                     /note="'Primer TATAplus'"
BASE COUNT       14 a      5 c      7 g     19 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      21 CTGAAAATTTATAA 8


RESULT  18
AX099856
LOCUS       AX099856       45 bp    DNA             PAT       02-APR-2001
DEFINITION  Sequence 5 from Patent WO0120014.
ACCESSION   AX099856
VERSION     AX099856.1  GI:13538882
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Schouten,G.J., Vogels,R. and Opstelten,D.J.
  TITLE     Modified adenoviral vectors for use in gene therapy
  JOURNAL   Patent: WO 0120014-A 5 22-MAR-2001;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
     primer_bind     1. .45
                     /note="'Primer TATAmin'"
BASE COUNT       19 a      7 c      5 g     14 t
```

```
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      29 CTGAAAATTTATAA 42


RESULT  19
AX137975/c
LOCUS       AX137975       45 bp    DNA             PAT       30-MAY-2001
DEFINITION  Sequence 4 from Patent EP1083229.
ACCESSION   AX137975
VERSION     AX137975.1  GI:14274070
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   .
  TITLE     Modified adenoviral vectors for use in gene therapy
  JOURNAL   Patent: EP 1083229-A 4 14-MAR-2001;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
     primer_bind     1. .45
                     /note="'Primer TATAplus'"
BASE COUNT       14 a      5 c      7 g     19 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      21 CTGAAAATTTATAA 8


RESULT  20
AX137976
LOCUS       AX137976       45 bp    DNA             PAT       30-MAY-2001
DEFINITION  Sequence 5 from Patent EP1083229.
ACCESSION   AX137976
VERSION     AX137976.1  GI:14274071
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   .
  TITLE     Modified adenoviral vectors for use in gene therapy
  JOURNAL   Patent: EP 1083229-A 5 14-MAR-2001;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
     primer_bind     1. .45
                     /note="'Primer TATAmin'"
BASE COUNT       19 a      7 c      5 g     14 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      29 CTGAAAATTTATAA 42


RESULT  21
AX159856/c
LOCUS       AX159856       50 bp    DNA             PAT       22-JUN-2001
DEFINITION  Sequence 3184 from Patent WO0140521.
ACCESSION   AX159856
VERSION     AX159856.1  GI:14541187
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Shimkets,R.A. and Leach,M.
  TITLE     Nucleic acids containing single nucleotide polymorphisms and
            methods of use thereof
  JOURNAL   Patent: WO 0140521-A 3184 07-JUN-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    25. .26
                     /note="Nucleotide deleted between bases 25 and 26
                     Accession number cg43063075"
     misc_feature    26
                     /note="2 of 2 allelic variants (3183 is other entry)"
BASE COUNT       10 a      10 c      8 g     22 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 6;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     155 aaggaaaatataaa 168
           ||||||||||||||
Db      36 AAGGAAAATATAAA 23


RESULT  22
AR146953
LOCUS       AR146953       18 bp    DNA             PAT       08-AUG-2001
DEFINITION  Sequence 13 from patent US 6221361.
ACCESSION   AR146953
VERSION     AR146953.1  GI:15110756
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Cochran,M.D. and Junker,D.E.
  TITLE     Recombinant swinepox virus
  JOURNAL   Patent: US 6221361-A 13 24-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unknown"
BASE COUNT        8 a       1 c       2 g       7 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    231 aattgaattgtaa 243
          |||||||||||||
Db      5 AATTGAATTGTAA 17


RESULT  23
AR121058
LOCUS       AR121058       20 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 79 from patent US 6159694.
ACCESSION   AR121058
VERSION     AR121058.1  GI:14104634
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Karras,J.G.
  TITLE     Antisense modulation of stat3 expression
  JOURNAL   Patent: US 6159694-A 79 12-DEC-2000;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        9 a      3 c      3 g      5 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    173 atttagataaaag 185
          |||||||||||||
Db      2 ATTTAGATAAAAG 14


RESULT  24
AX076045/c
LOCUS       AX076045       20 bp    DNA             PAT       06-FEB-2001
DEFINITION  Sequence 21 from Patent WO0104358.
ACCESSION   AX076045
VERSION     AX076045.1  GI:12710698
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Stuyver,L., Maertens,G. and van Geyt,C.
  TITLE     Detection of anti-hepatitis b drug resistance
  JOURNAL   Patent: WO 0104358-A 21 18-JAN-2001;
            INNOGENETICS N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="Hepatitis B virus"
                     /db_xref="taxon:10407"
BASE COUNT        6 a      3 c      4 g      7 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     72 tatatagctgaaa 84
          |||||||||||||
Db     15 TATATAGCTGAAA 3


RESULT  25
A97479
LOCUS       A97479         23 bp    DNA             PAT       26-JAN-2000
DEFINITION  Sequence 35 from Patent WO9916780.
ACCESSION   A97479
VERSION     A97479.1  GI:6780825
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Gala,J. and Vannuffel,P.
  TITLE     GENETIC SEQUENCES, DIAGNOSTIC AND/OR QUANTIFICATION METHODS AND
            DEVICES FOR THE IDENTIFICATION OF STAPHYLOCOCCI STRAINS
  JOURNAL   Patent: WO 9916780-A 35 08-APR-1999;
            GALA JEAN LUC (BE); UNIV LOUVAIN (BE)
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       11 a      2 c      3 g      7 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     58 aaagatatattaa 70
          |||||||||||||
Db      6 AAAGATATATTAA 18


RESULT  26
AX093544
LOCUS       AX093544       24 bp    DNA             PAT       30-MAR-2001
DEFINITION  Sequence 74 from Patent WO0118198.
ACCESSION   AX093544
VERSION     AX093544.1  GI:13509982
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Weissenbach,J. and Hazan,J.
  TITLE     Cloning, expression and characterisation of the spg4 gene
            responsible for the most frequent form of autosomal spastic
            paraplegia
  JOURNAL   Patent: WO 0118198-A 74 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Site accepteur d' pissage du g ne SPG4."
BASE COUNT        8 a      2 c      4 g     10 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    199 tttttattttaaa 211
          |||||||||||||
Db      3 TTTTTATTTTAAA 15


RESULT  27
AX164353
LOCUS       AX164353       24 bp    DNA             PAT       22-JUN-2001
DEFINITION  Sequence 183 from Patent WO0138564.
ACCESSION   AX164353
VERSION     AX164353.1  GI:14545287
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
```

```
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Rouleau,G.A., Lafreniere,R.G., Rochefort,D., Cossette,P. and
            Ragsdale,D.
  TITLE     Loci for idiopathic generalized epilepsy, mutations thereof and
            method using same to assess, diagnose, prognose or treat epilepsy
  JOURNAL   Patent: WO 0138564-A 183 31-MAY-2001;
            McGill University (CA)
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="synthetic oligonucleotide"
BASE COUNT        2 a      5 c      3 g     14 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    301 tttatgttttcat 313
          |||||||||||||
Db      8 TTTATGTTTTCAT 20


RESULT  28
AB069100/c
LOCUS       AB069100       24 bp    DNA             SYN       08-AUG-2001
DEFINITION  Synthetic construct DNA, forward primer for human STS sts-R133B1R
            at 1p36.
ACCESSION   AB069100
VERSION     AB069100.1  GI:15129904
KEYWORDS    .
SOURCE      synthetic construct DNA.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Chen,Y.Z., Hayashi,Y., Wu,J.G., Takaoka,E., Maekawa,K.,
            Watanabe,N., Inazawa,J., Hosoda,F., Arai,Y., Mizushima,H.,
            Morohashi,A., Ohira,M., Nakagawara,A., Liu,S., Hoshi,M., Horii,A.
            and Soeda,E.
  TITLE     A bac-based sts-content map spanning a 35-mb region of human
            chromosome 1p35-p36
  JOURNAL   Genomics. 74 (1), 55-70 (2001)
  MEDLINE   21269192
REFERENCE   2  (bases 1 to 24)
  AUTHORS   Horii,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2001) Akira Horii, Tohoku University School of
            Medicine, Molecular Pathology; 2-1 Seiryomachi, Aoba-ku, Sendai,
            Miyagi 980-8575, Japan (E-mail:horii@mail.cc.tohoku.ac.jp,
            Tel:81-22-717-8042, Fax:81-22-717-8047)
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     misc_feature    1. .24
                     /note="forward primer for human STS sts-R133B1R at 1p36
                     sts-R133B1R obtained from clones B367G16, B133B1, B247E16,
                     B123L12, B32A16, B76C16, B81K9, B32I16, Human BAC library
                     RPCI-11"
BASE COUNT        9 a      2 c      4 g      9 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 12;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    227 atataattgaatt 239
          |||||||||||||
Db     14 ATATAATTGAATT 2


RESULT  29
AX042574/c
LOCUS       AX042574       25 bp    DNA             PAT       23-NOV-2000
DEFINITION  Sequence 140 from Patent WO0065088.
ACCESSION   AX042574
VERSION     AX042574.1  GI:11341182
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Ulfendahl,P.J. and Wong,K.C.
  TITLE     Primers for identifying typing or classifying nucleic acids
  JOURNAL   Patent: WO 0065088-A 140 02-NOV-2000;
            Amersham Pharmacia Biotech AB (SE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="DQA1 Homozygote primer sequence"
BASE COUNT        4 a      5 c      3 g     13 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    237 attgtaaaaaaaa 249
          |||||||||||||
Db     17 ATTGTAAAAAAAA 5


RESULT  30
AX043268/c
LOCUS       AX043268       25 bp    DNA             PAT       23-NOV-2000
DEFINITION  Sequence 834 from Patent WO0065088.
ACCESSION   AX043268
VERSION     AX043268.1  GI:11341876
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Ulfendahl,P.J. and Wong,K.C.
  TITLE     Primers for identifying typing or classifying nucleic acids
  JOURNAL   Patent: WO 0065088-A 834 02-NOV-2000;
            Amersham Pharmacia Biotech AB (SE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="DQA1 Heterozygote Primer Sequence"
BASE COUNT        4 a      5 c      3 g     13 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    237 attgtaaaaaaaa 249
          |||||||||||||
Db     17 ATTGTAAAAAAAA 5


RESULT  31
AX039624
LOCUS       AX039624       26 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 13 from Patent WO0063441.
ACCESSION   AX039624
VERSION     AX039624.1  GI:11229653
```

```
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Herrnstadt,C. and Davis,R.E.
  TITLE     Single nucleotide polymorphisms in mitochondrial genes that segreg
            ate with alzheimer's disease
  JOURNAL   Patent: WO 0063441-A 13 26-OCT-2000;
            MITOKOR (US)
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
BASE COUNT        8 a      2 c      5 g     11 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    304 atgttttcatgtt 316
          |||||||||||||
Db      7 ATGTTTTCATGTT 19


RESULT  32
AX039654
LOCUS       AX039654       26 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 43 from Patent WO0063441.
ACCESSION   AX039654
VERSION     AX039654.1  GI:11229683
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Herrnstadt,C. and Davis,R.E.
  TITLE     Single nucleotide polymorphisms in mitochondrial genes that segreg
            ate with alzheimer's disease
  JOURNAL   Patent: WO 0063441-A 43 26-OCT-2000;
            MITOKOR (US)
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
BASE COUNT        8 a      2 c      5 g     11 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    304 atgttttcatgtt 316
          |||||||||||||
Db      7 ATGTTTTCATGTT 19


RESULT  33
AR014028
LOCUS       AR014028       28 bp    DNA             PAT       05-DEC-1998
DEFINITION  Sequence 64 from patent US 5773246.
ACCESSION   AR014028
VERSION     AR014028.1  GI:3971482
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Keene,J.D., Levine,T. and Gao,F.
  TITLE     Methods and compositions useful in the recognition, binding and
            expression of ribonucleic acids involved in cell growth, neoplasia
            and immunoregulation
  JOURNAL   Patent: US 5773246-A 64 30-JUN-1998;
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="unknown"
BASE COUNT        6 a      3 c      1 g     18 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttt 208
          |||||||||||||
Db      3 TTATTTTTATTTT 15


RESULT  34
I21978
LOCUS       I21978         28 bp    DNA             PAT       07-OCT-1996
DEFINITION  Sequence 64 from patent US 5525495.
ACCESSION   I21978
VERSION     I21978.1  GI:1602332
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Keene,J.D., Levine,T. and Gao,F.
  TITLE     Methods and compositions useful in the recognition, binding and
            expression of ribonucleic acids involved in cell growth, neoplasia
            and immunoregulation
  JOURNAL   Patent: US 5525495-A 64 11-JUN-1996;
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="unknown"
BASE COUNT        6 a      3 c      1 g     18 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttt 208
          |||||||||||||
Db      3 TTATTTTTATTTT 15


RESULT  35
AX012366
LOCUS       AX012366       29 bp    DNA             PAT       06-SEP-2000
DEFINITION  Sequence 26 from Patent EP0955369.
ACCESSION   AX012366
VERSION     AX012366.1  GI:9998412
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Burtscher,H.D., Mueller,R.D., Hoelke,W.D. and Millan,J.L.
  TITLE     High active alkaline phosphatase
  JOURNAL   Patent: EP 0955369-A 26 10-NOV-1999;
            ROCHE DIAGNOSTICS GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Artificial"
```

```
BASE COUNT        8 a      7 c      9 g      5 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      22 gcacagaagaatg 34
           |||||||||||||
Db       9 GCACAGAAGAATG 21


RESULT  36
E59972
LOCUS       E59972          29 bp    DNA             PAT       07-FEB-2001
DEFINITION  Highly active alkaline phosphatase.
ACCESSION   E59972
VERSION     E59972.1  GI:13017742
KEYWORDS    JP 1999332586-A/23.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Werner,H.R.M.M. and Burutosha,J.L.M.M.
  TITLE     Highly active alkaline phosphatase
  JOURNAL   Patent: JP 1999332586-A 23 07-DEC-1999;
            ROCHE DIAGNOSTICS GMBH
COMMENT     OS   Artificial Sequence
            PN   JP 1999332586-A/23
            PD   07-DEC-1999
            PF   06-MAY-1999 JP 1999126494
            PR   05-MAY-1998 DE   19819962:7
            PI   WERNER HOERUKU,REINA MULLER,HERUMUTTO BURUTOSHA, PI   JOSE
            LOUIS MILAN
            PC   C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/16, PC
            C12N15/00,C12N5/00
            CC
            FH   Key             Location/Qualifiers
            FT   source          1. .29
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      7 c      9 g      5 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      22 gcacagaagaatg 34
           |||||||||||||
Db       9 GCACAGAAGAATG 21


RESULT  37
AR028230
LOCUS       AR028230        30 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 79 from patent US 5858661.
ACCESSION   AR028230
VERSION     AR028230.1  GI:5940203
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Ataxia-telangiectasia gene and its genomic organization
  JOURNAL   Patent: US 5858661-A 79 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT       13 a      2 c      1 g     14 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     272 aaaaaattatttc 284
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT  38
AR138633
LOCUS       AR138633        30 bp    DNA             PAT       16-JUN-2001
DEFINITION  Sequence 158 from patent US 6200749.
ACCESSION   AR138633
VERSION     AR138633.1  GI:14480978
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Mutated forms of the ataxia-telangiectasia gene and method to
            screen for a partial A-T phenotype
  JOURNAL   Patent: US 6200749-A 158 13-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT       13 a      2 c      1 g     14 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     272 aaaaaattatttc 284
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT  39
AX063379/c
LOCUS       AX063379        30 bp    DNA             PAT       24-JAN-2001
DEFINITION  Sequence 42 from Patent WO0079009.
ACCESSION   AX063379
VERSION     AX063379.1  GI:12541169
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Nazarenko,I. and Rashtchian,A.
  TITLE     Improved primers and methods for the detection and discrimination
            of nucleic acids
  JOURNAL   Patent: WO 0079009-A 42 28-DEC-2000;
            Life Technologies, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligonucleotide"
BASE COUNT       10 a      4 c      7 g      9 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
```

```
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     40 atgaaataaagat 52
          |||||||||||||
Db     14 ATGAAATAAAGAT 2


RESULT  40
I95122/c
LOCUS       I95122          31 bp    DNA             PAT       01-DEC-1998
DEFINITION  Sequence 5 from patent US 5731425.
ACCESSION   I95122
VERSION     I95122.1  GI:3939592
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 31)
  AUTHORS   Brizzard,B.L., Bianca,D.W., Chubet,R.G., Vizard,D.L. and
            Hopp,T.Patrick.
  TITLE     Polypeptide surface marker for cells
  JOURNAL   Patent: US 5731425-A 5 24-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .31
                     /organism="unknown"
BASE COUNT       12 a     11 c      7 g      1 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    124 gtgttctcggggg 136
          |||||||||||||
Db     28 GTGTTCTCGGGGG 16


RESULT  41
XELARSE59
LOCUS       XELARSE59       33 bp    DNA             VRT       28-APR-1993
DEFINITION  Xenopus laevis autonomous replication sequence e59.
ACCESSION   K01606
VERSION     K01606.1  GI:213953
KEYWORDS    autonomous replication; mutational analysis.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Kearsey,S.
  TITLE     Structural requirements for the function of a yeast chromosomal
            replicator
  JOURNAL   Cell 37, 299-307 (1984)
  MEDLINE   84205653
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
BASE COUNT        8 a      4 c      2 g     19 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 5;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    298 atttttatgtttt 310
          |||||||||||||
Db      1 ATTTTTATGTTTT 13


RESULT  42
A41027/c
LOCUS       A41027          36 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 14 from Patent EP0626449.
ACCESSION   A41027
VERSION     A41027.1  GI:2296916
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Schreier,P.H., Stenzel,K.D., Adam,G.P. and Maiss,E.D.
  TITLE     DNA coding for plant virus sequences
  JOURNAL   Patent: EP 0626449-A 14 30-NOV-1994;
            BAYER AG (DE)
COMMENT     Other publication JP 6343469 941220
            Other publication CA 2124272 941129
            Other publication CN 1098744 950215
            Other publication AU 6191494 941201
            Other publication DE 4317845 941201
            Other publication ZA 9403730 950202.
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        9 a      5 c      5 g     17 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    156 aggaaaatataaa 168
          |||||||||||||
Db     24 AGGAAAATATAAA 12


RESULT  43
AR082586/c
LOCUS       AR082586        36 bp    DNA             PAT       31-AUG-2000
DEFINITION  Sequence 14 from patent US 5973135.
ACCESSION   AR082586
VERSION     AR082586.1  GI:10009308
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Schreier,P.Helmut, Stenzel,K., Adam,G. and Maiss,E.
  TITLE     DNA comprising plum pox virus and tomato spotted wilt virus cDNAS
            for disease resistance
  JOURNAL   Patent: US 5973135-A 14 26-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="unknown"
BASE COUNT        9 a      5 c      5 g     17 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    156 aggaaaatataaa 168
          |||||||||||||
Db     24 AGGAAAATATAAA 12


RESULT  44
AX167671
LOCUS       AX167671        36 bp    DNA             PAT       03-JUL-2001
DEFINITION  Sequence 16 from Patent WO0144277.
```

```
ACCESSION   AX167671
VERSION     AX167671.1  GI:14597058
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   wegrich Glover,L., Budziszewski,G.J., Levin,J.Z. and Zhou,Q.
  TITLE     Herbicide target genes and methods
  JOURNAL   Patent: WO 0144277-A 16 21-JUN-2001;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide"
BASE COUNT        6 a      4 c     11 g     15 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    303 tatgttttcatgt 315
          |||||||||||||
Db     21 TATGTTTTCATGT 33


RESULT  45
I28261/c
LOCUS       I28261          36 bp    DNA             PAT       06-FEB-1997
DEFINITION  Sequence 14 from patent US 5569823.
ACCESSION   I28261
VERSION     I28261.1  GI:1819037
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Schreier,P.H., Stenzel,K., Adam,G.unter and Maiss,E.
  TITLE     DNA comprising plum pox virus and tomato spotted wilt virus cDNAS
            for disease resistance
  JOURNAL   Patent: US 5569823-A 14 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="unknown"
BASE COUNT        9 a      5 c      5 g     17 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    156 aggaaaatataaa 168
          |||||||||||||
Db     24 AGGAAAATATAAA 12


Search completed: January 24, 2002, 03:22:36
Job time: 3813 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 01:12:27 ; Search time 1502.85 Seconds
(without alignments)
3589.565 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		327	100.0	327	6	AX004614	AX004614 Sequence
2		327	100.0	1392	1	L77965	L77965 Clostridium
3		327	100.0	1392	6	AX004613	AX004613 Sequence
4		199	60.9	54310	1	AP003515	AP003515 Clostridi
5		81.8	25.0	12900	3	AE001429	AE001429 Plasmodiu
6		79	24.2	39347	9	AL135906	AL135906 Human DNA
7		77.8	23.8	53932	2	AC023371	AC023371 Homo sapi
8		77.4	23.7	163443	2	AC006280	AC006280 Plasmodiu
9		77.4	23.7	205429	2	AC005506	AC005506 Plasmodiu
10	c	76.2	23.3	3392	3	AF300334	AF300334 Dictyoste
11	c	76.2	23.3	242513	2	AC079314	AC079314 Homo sapi
12	c	75.8	23.2	318221	2	PFMAL13P3	AL049184 Plasmodiu
13	c	75.6	23.1	140414	2	AF377947	AF377947 Oryza sat
14	c	75	22.9	178783	9	AC068139	AC068139 Homo sapi
15		74.4	22.8	156060	2	AC004153	AC004153 Plasmodiu
16	c	73.2	22.4	863	11	CNS06EVQ	AL395628 T7 end of
17		73	22.3	8622	8	YSCMTCYTOC	M97514 Saccharomyc
18	c	72.8	22.3	34119	8	AF222718	AF222718 Chrysodid
19		72.2	22.1	104992	2	AC005504	AC005504 Plasmodiu
20		72.2	22.1	162445	9	AL158151	AL158151 Human DNA
21		72.2	22.1	169546	2	AC004157	AC004157 Plasmodiu
22		72.2	22.1	199882	9	AL354720	AL354720 Human DNA
23		72	22.0	158398	2	AC011146	AC011146 Homo sapi
24	c	71.6	21.9	180903	9	AC073409	AC073409 Homo sapi
25		71.4	21.8	12029	3	AE001400	AE001400 Plasmodiu
26	c	71.4	21.8	175053	2	AC090014	AC090014 Homo sapi
27	c	71.2	21.8	110000	2	AL591074_2	Continuation (3 of
28		71	21.7	95477	9	AC007076	AC007076 Homo sapi
29	c	71	21.7	168799	9	AC009531	AC009531 Homo sapi
30	c	71	21.7	194038	9	AC010103	AC010103 Homo sapi
31	c	70.8	21.7	159475	2	AC021378	AC021378 Homo sapi
32		70.6	21.6	13433	3	AF315648	AF315648 Ceratitis
33	c	70.6	21.6	85779	8	SCE011856	AJ011856 Saccharom
34	c	70.6	21.6	122747	2	AC093220	AC093220 Homo sapi
35		70.6	21.6	159255	9	AF212831	AF212831 Homo sapi
36		70.6	21.6	161230	2	AC011355	AC011355 Homo sapi
37		70.6	21.6	234112	3	PFMAL4P2	AL035475 Plasmodiu
38		70.6	21.6	340000	9	HS21C013	AL163213 Homo sapi
39	c	70.4	21.5	155456	2	AC027753	AC027753 Homo sapi
40	c	70.4	21.5	160624	9	AC060835	AC060835 Homo sapi
41		70.4	21.5	172758	2	AC022553	AC022553 Homo sapi
42	c	70.4	21.5	199551	2	AC006281	AC006281 Plasmodiu
43	c	70.2	21.5	137342	9	AL392048	AL392048 Human DNA
44	c	70.2	21.5	180388	9	HUMRETBLAS	L11910 Human retin
45	c	70.2	21.5	183584	9	AC012492	AC012492 Homo sapi

ALIGNMENTS

RESULT 1
AX004614
LOCUS AX004614 327 bp DNA PAT 24-AUG-2000
DEFINITION Sequence 2 from Patent WO9915669.
ACCESSION AX004614
VERSION AX004614.1 GI:9928055
KEYWORDS .
SOURCE Clostridium perfringens.
ORGANISM Clostridium perfringens
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE 1 (bases 1 to 327)
AUTHORS Gibert,M. and Popoff,M.R.
TITLE Clostridium toxin, and method for preparing immunogenic compositions
JOURNAL Patent: WO 9915669-A 2 01-APR-1999;
GIBERT MARYSE (FR); PASTEUR INSTITUT (FR)
FEATURES Location/Qualifiers
source 1. .327
/organism="Clostridium perfringens"

```
                     /db_xref="taxon:1502"
BASE COUNT      141 a     13 c     44 g    129 t
ORIGIN


  Query Match             100.0%;  Score 327;  DB 6;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-28;
  Matches  327;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATTTGGGATATCTTAAATTTAGCACAGAAGAATGTTTAAATGAAATAAAGATAATAAAAA 60

Qy     61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATATATTAATTATATAGCTGAAAATTTATAATTATATGATAAGTATAGTTAATAAATAA 120

Qy    121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAGTGTTCTCGGGGGACACTTTTTTGTTTTAAAAAGGAAAATATAAATAAAATTTAGAT 180

Qy    181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAAGTGTAAAATAATTATTTTTATTTTAAATTTGTTAAAAATTTGATATAATTGAATTG 240

Qy    241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TAAAAAAAATTTCAGGGGGGAATATAAATGAAAAAAATTATTTCAAAGTTTACTGTAATT 300

Qy    301 tttatgttttcatgttttcttattgtt 327
          |||||||||||||||||||||||||||
Db    301 TTTATGTTTTCATGTTTTCTTATTGTT 327


RESULT   2
L77965
LOCUS       L77965       1392 bp    DNA             BCT       28-JUL-1998
DEFINITION  Clostridium perfringens C beta 2 toxin gene, complete cds.
ACCESSION   L77965
VERSION     L77965.1  GI:3342214
KEYWORDS    .
SOURCE      Clostridium perfringens C.
  ORGANISM  Clostridium perfringens C
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 1392)
  AUTHORS   Gibert,M., Jolivet-Reynaud,C. and Popoff,M.R.
  TITLE     Beta2 toxin, a novel toxin produced by Clostridium perfringens
  JOURNAL   Gene 203 (1), 65-73 (1997)
  MEDLINE   98085977
  REMARK    Erratum:[[published erratum appears in Gene 1998 Mar
            27;210(1):173]]
REFERENCE   2  (bases 1 to 1392)
  AUTHORS   Popoff,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1998) Toxines Microbiennes, Institut Pasteur,
            Paris cedex 15 75724, France
COMMENT     GSDB:S:76036.
            [Flatfile retrieved from GSDB Fri Jul 24 15:39:17 1998].
FEATURES             Location/Qualifiers
     source          1. .1392
                     /organism="Clostridium perfringens C"
                     /strain="CWC245"
                     /db_xref="taxon:79668"
     RBS             255. .260
                     /note="putative"
     sig_peptide     268. .357
                     /note="putative"
     CDS             268. .1065
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="beta 2 toxin"
                     /protein_id="AAC27654.1"
                     /db_xref="GI:3342215"
                     /translation="MKKIISKFTVIFMFSCFLIVGAISPMKASAKEIDAYRKVMENYL
                     NALKNYDINTVVNISEDERVNNVEQYREMLEDFKYDPNQQLKSFEILNSQKSDNKEIF
                     NVKTEFLNGAIYDMEFTVSSKDGKLIVSDMERTKVENEGKYILTPSFRTQVCTWDDEL
                     AQAIGGVYPQTYSDRFTYYADNILLNFRQYATSGSRDLKVEYSVVDHWMWKDDVKASQ
                     MVYGQNPDSARQIRLYIEKGQSFYKYRIRIKNFTPASIRVFGEGYCA"
     mat_peptide     358. .1062
                     /note="putative"
                     /product="beta 2 toxin"
     terminator      1066. .1104
                     /note="putative"
BASE COUNT      606 a    115 c    209 g    462 t
ORIGIN


  Query Match             100.0%;  Score 327;  DB 1;  Length 1392;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-28;
  Matches  327;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATTTGGGATATCTTAAATTTAGCACAGAAGAATGTTTAAATGAAATAAAGATAATAAAAA 60

Qy     61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATATATTAATTATATAGCTGAAAATTTATAATTATATGATAAGTATAGTTAATAAATAA 120

Qy    121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAGTGTTCTCGGGGGACACTTTTTTGTTTTAAAAAGGAAAATATAAATAAAATTTAGAT 180

Qy    181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAAGTGTAAAATAATTATTTTTATTTTAAATTTGTTAAAAATTTGATATAATTGAATTG 240

Qy    241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TAAAAAAAATTTCAGGGGGGAATATAAATGAAAAAAATTATTTCAAAGTTTACTGTAATT 300

Qy    301 tttatgttttcatgttttcttattgtt 327
          |||||||||||||||||||||||||||
Db    301 TTTATGTTTTCATGTTTTCTTATTGTT 327


RESULT   3
AX004613
LOCUS       AX004613     1392 bp    DNA             PAT       24-AUG-2000
DEFINITION  Sequence 1 from Patent WO9915669.
ACCESSION   AX004613
VERSION     AX004613.1  GI:9928053
KEYWORDS    .
SOURCE      Clostridium perfringens.
  ORGANISM  Clostridium perfringens
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 1392)
  AUTHORS   Gibert,M. and Popoff,M.R.
  TITLE     Clostridium toxin, and method for preparing immunogenic
            compositions
  JOURNAL   Patent: WO 9915669-A 1 01-APR-1999;
            GIBERT MARYSE (FR); PASTEUR INSTITUT (FR)
FEATURES             Location/Qualifiers
     source          1. .1392
                     /organism="Clostridium perfringens"
                     /db_xref="taxon:1502"
     CDS             268. .1065
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC04901.1"
                     /db_xref="GI:9928054"
                     /translation="MKKIISKFTVIFMFSCFLIVGAISPMKASAKEIDAYRKVMENYL
```

```
                    NALKNYDINTVVNISEDERVNNVEQYREMLEDFKYDPNQQLKSFEILNSQKSDNKEIF
                    NVKTEFLNGAIYDMEFTVSSKDGKLIVSDMERTKVENEGKYILTPSFRTQVCTWDDEL
                    AQAIGGVYPQTYSDRFTYYADNILLNFRQYATSGSRDLKVEYSVVDHWMWKDDVKASQ
                    MVYGQNPDSARQIRLYIEKGQSFYKYRIRIKNFTPASIRVFGEGYCA"
BASE COUNT      606 a    115 c    209 g    462 t
ORIGIN


  Query Match           100.0%;  Score 327;  DB 6;  Length 1392;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-28;
  Matches  327;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATTTGGGATATCTTAAATTTAGCACAGAAGAATGTTTAAATGAAATAAAGATAATAAAAA 60

Qy     61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATATATTAATTATATAGCTGAAAATTTATAATTATATGATAAGTATAGTTAATAAATAA 120

Qy    121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAGTGTTCTCGGGGGACACTTTTTTGTTTTAAAAAGGAAAATATAAATAAAATTTAGAT 180

Qy    181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAAGTGTAAAATAATTATTTTTATTTTAAATTTGTTAAAAATTTGATATAATTGAATTG 240

Qy    241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TAAAAAAAATTTCAGGGGGGAATATAAATGAAAAAAATTATTTCAAAGTTTACTGTAATT 300

Qy    301 tttatgttttcatgttttcttattgtt 327
          |||||||||||||||||||||||||||
Db    301 TTTATGTTTTCATGTTTTCTTATTGTT 327


RESULT  4
AP003515
LOCUS       AP003515    54310 bp    DNA   circular  BCT       10-AUG-2001
DEFINITION  Clostridium perfringens plasmid pCP13 DNA, complete sequence.
ACCESSION   AP003515
VERSION     AP003515.1  GI:15076712
KEYWORDS    .
SOURCE      Clostridium perfringens (strain:13) plasmid:pCP13 DNA.
  ORGANISM  Clostridium perfringens
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 54310)
  AUTHORS   Ohtani,K., Ohshima,S., Hirakawa,H., Ohshima,K., Shiba,T.,
            Shimizu,T., Hattori,M., Kuhara,S., Hayashi,H. and Shimizu,T.
  TITLE     Complete Nucleotide Sequence of the Virulence Plasmid pCP13 from
            Clostridium perfringens
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 54310)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2001) Tohru Shimizu, Institute of Basic Medical
            Sciences, University of Tukuba, Department of Microbiology; 1-1-1
            Tennohdai, Tsukuba, Ibaraki 305-8575, Japan
            (E-mail:tshimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
            Fax:81-298-53-3354)
FEATURES             Location/Qualifiers
     source          1. .54310
                     /organism="Clostridium perfringens"
                     /plasmid="pCP13"
                     /strain="13"
                     /db_xref="taxon:1502"
                     /note="anaerobic pathogen for gas gangrene"
     gene            940. .1692
                     /gene="soj"
     CDS             940. .1692
                     /gene="soj"
                     /note="250 aa, similar to pir:I40444 Spo0A activation
                     inhibitor soj from Bacillus subtilis (253 aa); 37%
                     identity in 250 aa overlap
                     PCP01
                     ParA family"
                     /codon_start=1
                     /transl_table=11
                     /product="Soj protein"
                     /protein_id="BAB62438.1"
                     /db_xref="GI:15076713"
                     /translation="MKKISVFNIKGGVAKTTSTANFGACLEEKGDRVLLVDLDPQSNL
                     TKLFKAYSMEDVSIADVLLDKNLDIEKVIKKTDFENIDILPSNVTLAFAERKILLDVN
                     RSQQNRLAKALEEIEDKYDYCLIDCPPALNMITVNALCASDEVLVPIKIDKFALDGLE
                     YLLDSIEEIKDEFNPNLNFKGCFITMDSSTTVNKVIKQELKSVLGEKMFNTSIHQNIK
                     VVESTFEECPVVFSSKKARASLNYKDLSKEIF"
     gene            1751. .3031
                     /gene="parB"
     CDS             1751. .3031
                     /gene="parB"
                     /note="426 aa, similar to gpu:AF300944_3 presumptive ParB
                     protein from Lactococcus lactis subsplactis (242 aa); 30%
                     identity in 266 aa overlap
                     PCP02"
                     /codon_start=1
                     /transl_table=11
                     /product="ParB protein"
                     /protein_id="BAB62439.1"
                     /db_xref="GI:15076714"
                     /translation="MAKKFSISEGMLNGISKNTRKVEEFQAKENFKIEYINIDRIKRN
                     EKNFYEIVDIESLAEDIKLNGLNHNLVVRKLDNDMYELISGERRYTALSKLVNEGNKE
                     FNLVPCKVIESNDIDSEIILIQANAQSRELTEVEKLTQVERLQEFYKIKKENGEKVPG
                     KIRDIIANDLNLSATQVGRYERINNKLIPELKAVIEQGNLTIANASEFSSLSEENQRV
                     ILSIIDDKTNMSKQEAVDLKNKLKKIEEEKELELKKAYEEKELELKRLEEEKKNQVSK
                     LKSENENLKKKLDSNNIEEERKEIEGQIKIEFEEKLKNEKVILEEELKSKYDKKIEDI
                     TKEAKENNLEKQRLKDELSKLKEKSNNEVDIKNTKENFVLVQNLKLIDNSFKNLKSQI
                     NKMKKENVKVAEETKAKEFLEKYQKEISDLLKKL"
     gene            3147. .3509
                     /gene="PCP03"
     CDS             3147. .3509
                     /gene="PCP03"
                     /note="120 aa, no significant homology"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB62440.1"
                     /db_xref="GI:15076715"
                     /translation="MEKILAEKRINISFYKRKNGALVTTLYLPPKWLEVIGITENERE
                     CFFYIEDKAIKISKEKQSEEAKEKTISFSKTSTKTYLNNKWLEYLGVSEDERSCIIEL
                     RKKDITLVKDNGRDILDI"
     gene            3773. .4024
                     /gene="PCP04"
     CDS             3773. .4024
                     /gene="PCP04"
                     /note="83 aa, similar to pir:T14710 probable transposase
                     from Yersinia pestis (402 aa); 44% identity in 50 aa
                     overlap
                     truncated"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transposase"
                     /protein_id="BAB62442.1"
                     /db_xref="GI:15076717"
                     /translation="MSKDFLRNYIKEQNFNDSNNVLNSLKDLFNDVLQEVVSAELDDM
                     LGYEKNLTNQFYNSRNGYSKKTIKSELGTITLNILRYRN"
     gene            4040. .4222
                     /gene="PCP05"
     CDS             4040. .4222
                     /gene="PCP05"
                     /note="60 aa, similar to gp:AF143819_1 transposase-like
                     protein from Escherichia coli (402 aa); 38% identity in 60
                     aa overlap
                     truncated"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="probable transposase"
                     /protein_id="BAB62443.1"
                     /db_xref="GI:15076718"
                     /translation="MISKHQRNINRIEDKNLNLYASGITTRDVAGQIKALYDIEISAE
                     TVSNITNRIMPLVSEW"
     gene            4588. .4746
                     /gene="PCP06"
     CDS             4588. .4746
                     /gene="PCP06"
                     /note="52 aa, similar to pir:T43600 probable transposase
                     from Yersinia pestis (105 aa); 56% identity in 50 aa
                     overlap
                     truncated"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transposase"
                     /protein_id="BAB62444.1"
                     /db_xref="GI:15076719"
                     /translation="MLYTTNVIESLNSQFRKFTKTKLIFPNDVSLIKMLYLATEKVNK
                     KWTRNYPN"
     gene            complement(5169. .5804)
                     /gene="PCP07"
     CDS             complement(5169. .5804)
                     /gene="PCP07"
                     /function="ATP-binding protein"
                     /note="211 aa, similar to gp:AP001508_4 ABC transporter
                     (ATP-binding protein) from Bacillus halodurans (213 aa);
                     49% identity in 214 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="probable ABC transporter"
                     /protein_id="BAB62445.1"
                     /db_xref="GI:15076720"
                     /translation="MNIIEISNLNKKYFDKVIFKDFSLSIKKGEMIAISGRSGCGKST
                     LLNMIGLIEKFDSGEIIIDGVKNIKINSKLANKFLREKISYLFQNFALVDEETVEENL
                     RLAIKHTIKNTKKIEEEIIRCLKFVGLEGCQKNYIYELSGGEQQRVAIARLMLKPSEI
                     ILADEPTGSLDEENRDIIISLLKELNESGKTIIIVTHDNYVAKQADRIIFL"
     gene            complement(5804. .7966)
                     /gene="PCP08"
     CDS             complement(5804. .7966)
                     /gene="PCP08"
                     /note="720 aa, similar to gp:AP001508_3 BH0280 gene
                     product from Bacillus halodurans (713 aa); 23% identity in
                     661 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="BAB62446.1"
                     /db_xref="GI:15076721"
                     /translation="MKKKVALIILFIIITILSFEGVYSVRNHTEFMKLKNLQGNLENF
                     EVTISIPDRENYNTAYESIIKSLDEYNGNIFFSEVDINEDVRKYINYGYFSNKETENH
                     IPIVSGRFFHEKDNIDSYLSTIDSEDTQQIGVINDFNGKNIHEIRTIKSKLDINTFEN
                     LFIVQIENEQILDKLIEDLKSESIIVQKRVMGDSSQYNTETLKIILVVCFIGLIFMIF
                     YQVLGSYKKIGIQKLLGHSTFVMLKERLLEVLRIEVIVMLVVTVLLVFFNFKTFNSLF
                     WKFMLELICIYSIMIIFTIVSVIIPYIYVSKITLSNIIKNKRPVKSIIILNSIVKVIL
                     ASIILIFFSNALDDLSSIGKGYEKNYKVWEETKQYYILPELGFNDESIQSFSIEEMEK
                     ERAVYLYFNKQGAILADFNRYEPTSMEEAKQMLPEEYMRETIIVNPNYLLKHKVYDVD
                     GNIINISEDEKDRILLVPEKYRNFEKEILEYYGYNSQEPCSSTTCSHKTADGKLNLVE
                     QKQKIIWMKSNQKYFSYLLDVNPEEGNYVTDPIVSVLTESNDKLVSYYKIIGYNNSPF
                     KIRANSEEEVINGLEKYYDMSVYLIDPYNLYDNVASTIINIKAKVKVIIFAIVILLAV
                     ISIIILQNTSLYFNQNKNKIIVKKLHGYRLIYRYMNYFIMVLITWTCPLAIASLITKD
                     INIIYFTLILVVIELVFIIFNINSLEKKNLIKVIKGEY"
     gene            complement(8043. .8354)
                     /gene="PCP09"
     CDS             complement(8043. .8354)
                     /gene="PCP09"
                     /note="103 aa, no significant homology"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB62447.1"
                     /db_xref="GI:15076722"
                     /translation="MKRNKLIKTFIASLLLVGSLGTTALAYTHSSDKFEAASLPGLFS
                     QAQSSKFYCGEQKHRATARVKVGTTLYEAKDIKDAKLTAHAQTKYYKGVTEWNSYYAH
                     L"
     gene            8779. .9012
                     /gene="PCP10"
     CDS             8779. .9012
                     /gene="PCP10"
                     /note="77 aa, similar to probable transposase from
                     Yersinia pestis plasmid pMT1 (402 aa); 25% identity in 158
                     aa overlap
                     truncated"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transposase"
                     /protein_id="BAB62448.1"
                     /db_xref="GI:15076723"
                     /translation="MKEVLSICMSTNEGSKFWLSPHKDKKEFAKDLKTIYGSVNETEG
                     MKNLIELREKWGSKYPNVVKSWKDNWDNLSTFF"
     gene            9127. .9366
                     /gene="PCP11"

  Query Match            60.9%;  Score 199;  DB 1;  Length 54310;
  Best Local Similarity  86.3%;  Pred. No. 2.2e-14;
  Matches  220;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      73 atatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgttctcg 132
           ||| |  |  || |      ||| | | ||| |||| |  |||||  || |||| |
Db   13459 ATAGAATTTTAACTCATGTTTTAAAAGTTAAATATATATTTTAAATTTAAGGTGTCCCAC 13518

Qy     133 ggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaa 192
           |||||||  |||||||||| |||||| |||||| ||||||||||||||||||||||||||
Db   13519 GGGGACATCTTTTTGTTTTTAAAAGGTAAATATGAATAAAATTTAGATAAAAGTGTAAAA 13578

Qy     193 taattatttttatttaaatttgttaaaaatttgatataattgaattgtaaaaaaaattt 252
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   13579 GAATTATTTTTATTTTAAATTTGTTAAAATTTTGATATAATTGAATTGTAAAAAAAATTT 13638

Qy     253 caggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttca 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13639 CAGGGGGGAATATAAATGAAAAAAATTATTTCAAAGTTTACTGTAATTTTTATGTTTTCA 13698

Qy     313 tgttttcttattgtt 327
           | ||||||||||||||
Db   13699 TATTTTCTTATTGTT 13713


RESULT   5
AE001429
LOCUS       AE001429    12900 bp    DNA             INV       06-NOV-1998
DEFINITION  Plasmodium falciparum chromosome 2, section 66 of 73 of the
            complete sequence.
ACCESSION   AE001429 AE001362
VERSION     AE001429.1  GI:3845321
KEYWORDS    .
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 12900)
  AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
            Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,
            Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M.,
            Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
            Smith,H.O., Fraser,C.M., Hoffman,S.L. et,al.
  TITLE     Chromosome 2 sequence of the human malaria parasite Plasmodium
            falciparum
  JOURNAL   Science 282 (5391), 1126-1132 (1998)
  MEDLINE   99021743
  REMARK    Erratum:[[published erratum appears in Science 1998 Dec
            4;282(5395):1827]]
REFERENCE   2  (bases 1 to 12900)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
```

```
           Medical Center Drive, Rockville, MD 20814, USA
FEATURES             Location/Qualifiers
     source          1. .12900
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="2"
     gene            11045. .12083
                     /gene="PFB0955w"
     CDS             join(11045. .11113,11229. .12083)
                     /gene="PFB0955w"
                     /note="identified by sequence similarity; putative"
                     /codon_start=1
                     /product="rifin"
                     /protein_id="AAC71980.1"
                     /db_xref="GI:3845322"
                     /translation="MNIYYINMLVMSILLIVLFLSYNVNNHNKKYNVGYIQNNRQMIM
                     MKSRRLAEIQLPKCPHYNNDPELKKIIDKLNEERIKKYIETNNSFEELHGLLVKERTK
                     SLYENGMKKSSNMEKELLKKYDDSIRDEHNVISKSGIYTSDYRKLYDKSCDYQNQKIL
                     RDELASCCKVHDNYLDNLKKGCFGGVGICTLCSLLVSNIGIGYAVTAAKEVITGLYSL
                     DIANKFTKALAGIYFFFSSSIENAGVSGVTIFYWDSMRMASIASSTINPYGIAALVLI
                     VLVVVLIVLYIWLYRRRKKSWKHECKKHLST"
BASE COUNT     5356 a    845 c   1167 g   5532 t
ORIGIN


  Query Match           25.0%;  Score 81.8;  DB 3;  Length 12900;
  Best Local Similarity  54.0%;  Pred. No. 0.25;
  Matches  167;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy       2 tttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaag 61
           |||  | | |||  | | |   | |  ||    || || |   ||    || |   ||
Db    6596 TTTCCGTTTTCTCTATTATTATATACTAGGGGTTTAAATTATTATTTTTATTAATGAATT 6655

Qy      62 atatattaattatatagctgaaaatttataattatatgataagtatagttaataaataaa 121
           ||||||   |   |   | ||||| ||||  |||||  | | |||   |||  |   ||
Db    6656 CTATATTTTATTATTTATTAAAAATATATACATATATATTTAATATGTCTAAACATATAA 6715

Qy     122 aagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagata 181
           |    | |      | |  | ||| || ||   |  |||||| ||||  ||||  | |
Db    6716 TAAATATATTTATATATATATATTTTTTATATTTATTAAAATATGAATATCATTTTTAGA 6775

Qy     182 aaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattgt 241
           | |      || |  || ||||| ||| |    | | |||||| |  ||||||  || | |
Db    6776 ATAAAAAGAATTTTTTTTTTTTTTTTTAATTAATTTAAAAAATATATTATAATAAAACTTT 6835

Qy     242 aaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaattt 301
           | ||| |    |      |||  |||||||||||| |||| |||   |  | |||  |
Db    6836 TATAAATAAAAAAATTCTAAATTAAAATGAAAAAAATAATTTTAAACAATTTTATAAAAT 6895

Qy     302 ttatgtttt 310
           || | ||
Db    6896 GAATATGTT 6904


RESULT   6
AL135906
LOCUS       AL135906    39347 bp    DNA             PRI       26-APR-2000
DEFINITION  Human DNA sequence from clone RP1-20N11 on chromosome 6 Contains
            GSSs, complete sequence.
ACCESSION   AL135906
VERSION     AL135906.19  GI:7248200
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39347)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 15, 2000 this sequence version replaced gi:7210154.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-20N11 is from the library RPCI-1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP1-20N11 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP1-20N11 is at 39347 in this sequence.
            The true right end of clone RP3-482L3 is at 103 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .39347
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP1-20N11"
                     /clone_lib="RPCI-1"
     repeat_region   2. .1136
                     /note="L1PA12 repeat: matches -1421. .-291 of consensus"
     repeat_region   3460. .3893
                     /note="L1M3c repeat: matches 999. .1402 of consensus"
     repeat_region   3894. .4213
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   4214. .4415
                     /note="L1M3c repeat: matches 1402. .1610 of consensus"
     repeat_region   4426. .4777
                     /note="MER47A repeat: matches 13. .366 of consensus"
     repeat_region   4779. .5831
                     /note="L1M2 repeat: matches 1528. .2663 of consensus"
     repeat_region   6028. .6073
                     /note="L1 repeat: matches 2729. .2777 of consensus"
     repeat_region   6091. .8167
                     /note="L1PA15 repeat: matches 4085. .6157 of consensus"
     repeat_region   8287. .8567
                     /note="AluSx repeat: matches 3. .284 of consensus"
     repeat_region   8792. .8965
                     /note="L1 repeat: matches 2737. .2901 of consensus"
     repeat_region   8959. .9103
                     /note="L1M3c repeat: matches 1092. .1236 of consensus"
     repeat_region   9227. .10104
                     /note="L1 repeat: matches 2903. .3922 of consensus"
     repeat_region   10091. .10657
                     /note="L1M2 repeat: matches 5115. .5687 of consensus"
     repeat_region   10657. .11436
                     /note="L1PA2 repeat: matches 3. .776 of consensus"
     repeat_region   11432. .16672
                     /note="L1PA2 repeat: matches 900. .6146 of consensus"
     repeat_region   16682. .16795
                     /note="L1P3 repeat: matches 5677. .5791 of consensus"
     repeat_region   16821. .16860
                     /note="20 copies 2 mer ta 90% conserved"
     repeat_region   16886. .17035
                     /note="MER4A2 repeat: matches 2. .152 of consensus"
```

```
     repeat_region   17036. .17343
                     /note="MER4A repeat: matches 367. .646 of consensus"
     repeat_region   17377. .17678
                     /note="L1MA9 repeat: matches 5783. .6112 of consensus"
     repeat_region   17679. .17996
                     /note="AluYb8 repeat: matches 1. .318 of consensus"
     repeat_region   17997. .18193
                     /note="L1MA9 repeat: matches 6112. .6300 of consensus"
     repeat_region   19452. .20219
                     /note="L1ME repeat: matches 429. .1193 of consensus"
     misc_feature    complement(20068. .20528)
                     /note="match: GSS: Em:AQ218253"
     repeat_region   23091. .23131
                     /note="MLT1I repeat: matches 370. .409 of consensus"
     repeat_region   23144. .23291
                     /note="L1PA4 repeat: matches 5999. .6146 of consensus"
     repeat_region   25530. .25553
                     /note="12 copies 2 mer ca 95% conserved"
     repeat_region   25570. .25754
                     /note="AluJo/FRAM repeat: matches 173. .291 of consensus"
     repeat_region   25827. .26622
                     /note="398 copies 2 mer aa 58% conserved"
     misc_feature    25879. .26639
                     /note="Tandem repeat. Tandem repeat contains forced join.
                     BamHI and EcoRI bot suggest approx 200bp missing."
     misc_feature    complement(26230. .26730)
                     /note="match: GSS: Em:B69019"
     misc_feature    complement(26267. .26664)
                     /note="match: GSS: Em:AQ230058"
     repeat_region   27019. .27243
                     /note="LTR40a repeat: matches 282. .516 of consensus"
     repeat_region   27306. .28098
                     /note="L1ME2 repeat: matches 5094. .5922 of consensus"
     repeat_region   28440. .28705
                     /note="133 copies 2 mer aa 54% conserved"
     repeat_region   29274. .29834
                     /note="L1MB6 repeat: matches 5552. .6128 of consensus"
     repeat_region   29885. .31038
                     /note="L1M4 repeat: matches 862. .2044 of consensus"
     repeat_region   31742. .32332
                     /note="L1MD1 repeat: matches 5317. .5893 of consensus"
     misc_feature    complement(32119. .32398)
                     /note="match: GSS: Em:AQ078965"
     repeat_region   32769. .32872
                     /note="52 copies 2 mer tt 64% conserved"
     misc_feature    complement(32852. .33057)
                     /note="match: GSS: Em:AQ780759"
     repeat_region   33046. .33127
                     /note="41 copies 2 mer tt 73% conserved"
     repeat_region   33143. .34362
                     /note="L1PA7 repeat: matches 4855. .6127 of consensus"
     repeat_region   34564. .34625
                     /note="31 copies 2 mer ta 71% conserved"
     repeat_region   34637. .34664
                     /note="14 copies 2 mer ac 89% conserved"
     repeat_region   34745. .36725
                     /note="L1PA13 repeat: matches 2266. .4198 of consensus"
     repeat_region   36726. .37032
                     /note="AluY repeat: matches 3. .308 of consensus"
     repeat_region   37033. .38914
                     /note="L1PA13 repeat: matches 4198. .6145 of consensus"
     repeat_region   39090. .39341
                     /note="TIGGER2 repeat: matches 2448. .2717 of consensus"
BASE COUNT    14429 a   7155 c   6738 g  11025 t
ORIGIN


  Query Match           24.2%;  Score 79;  DB 9;  Length 39347;
  Best Local Similarity  54.7%;  Pred. No. 0.43;
  Matches  157;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy     24 acagaagaatgtttaaatgaaataaagataataaaaagatatattaattatatagctgaa 83
          | | || |    | ||||  | || | |||  ||| | |||||| || ||||||    |
```

```
Db  25934 ATAAAATATATATAAAATATATTATATATATAAAATATATATATAAAATATATATAAAAT 25993

Qy     84 aatttataattatatgataagtatagttaataaataaaaagtgttctcgggggacacttt 143
          |  |||||  ||||  |||  ||||    ||| ||||||  | || |      | | | |
Db  25994 ATATTATATATATAAAATATATATAAAATATATATAAAATATATTATATATATAAAATAT 26053

Qy    144 tttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaataattatttttt 203
           |       | |  ||||||||    | || || |  | |    |||   ||||| | |
Db  26054 ATATAAAATATATATAAAATATATTATATATATAAAATATATATAAAATATATTATATAT 26113

Qy    204 attttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaat 263
          ||   | || | | ||| || | |||||  | || | || | ||||| |         ||
Db  26114 ATAAAATATATATAAAATATATTATATATATAAAATATATATAAAATATATATAAAATAT 26173

Qy    264 ataaatgaaaaaaattatttcaaagtttactgtaatttttatgtttt 310
          || |   | | ||| ||| |  ||  |   | |||  | |||  | |
Db  26174 ATTATATATATAAAATATATATAAAATATATATAAAATATATTATAT 26220


RESULT   7
AC023371
LOCUS       AC023371    53932 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-21D18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023371
VERSION     AC023371.2  GI:9123990
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 53932)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21D18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 53932)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6970502.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3985
                Center clone name: 21_D_18
            --------------.
            * NOTE: This record contains 55 individual
```

```
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      869: contig of 869 bp in length
*      870 969: gap of      100 bp
*      970     1815: contig of 846 bp in length
*     1816 1915: gap of      100 bp
*     1916     2797: contig of 882 bp in length
*     2798 2897: gap of      100 bp
*     2898     3768: contig of 871 bp in length
*     3769 3868: gap of      100 bp
*     3869     4769: contig of 901 bp in length
*     4770 4869: gap of      100 bp
*     4870     5739: contig of 870 bp in length
*     5740 5839: gap of      100 bp
*     5840     6703: contig of 864 bp in length
*     6704 6803: gap of      100 bp
*     6804     7675: contig of 872 bp in length
*     7676 7775: gap of      100 bp
*     7776     8671: contig of 896 bp in length
*     8672 8771: gap of      100 bp
*     8772     9639: contig of 868 bp in length
*     9640 9739: gap of      100 bp
*     9740    10613: contig of 874 bp in length
*    10614 10713: gap of      100 bp
*    10714    11572: contig of 859 bp in length
*    11573 11672: gap of      100 bp
*    11673    12528: contig of 856 bp in length
*    12529 12628: gap of      100 bp
*    12629    13488: contig of 860 bp in length
*    13489 13588: gap of      100 bp
*    13589    14469: contig of 881 bp in length
*    14470 14569: gap of      100 bp
*    14570    15452: contig of 883 bp in length
*    15453 15552: gap of      100 bp
*    15553    16429: contig of 877 bp in length
*    16430 16529: gap of      100 bp
*    16530    17401: contig of 872 bp in length
*    17402 17501: gap of      100 bp
*    17502    18387: contig of 886 bp in length
*    18388 18487: gap of      100 bp
*    18488    19361: contig of 874 bp in length
*    19362 19461: gap of      100 bp
*    19462    20326: contig of 865 bp in length
*    20327 20426: gap of      100 bp
*    20427    21291: contig of 865 bp in length
*    21292 21391: gap of      100 bp
*    21392    22281: contig of 890 bp in length
*    22282 22381: gap of      100 bp
*    22382    23271: contig of 890 bp in length
*    23272 23371: gap of      100 bp
*    23372    24249: contig of 878 bp in length
*    24250 24349: gap of      100 bp
*    24350    25213: contig of 864 bp in length
*    25214 25313: gap of      100 bp
*    25314    26162: contig of 849 bp in length
*    26163 26262: gap of      100 bp
*    26263    27122: contig of 860 bp in length
*    27123 27222: gap of      100 bp
*    27223    28086: contig of 864 bp in length
*    28087 28186: gap of      100 bp
*    28187    29073: contig of 887 bp in length
*    29074 29173: gap of      100 bp
*    29174    30055: contig of 882 bp in length
*    30056 30155: gap of      100 bp
*    30156    31136: contig of 981 bp in length
*    31137 31236: gap of      100 bp
*    31237    32116: contig of 880 bp in length
*    32117 32216: gap of      100 bp
*    32217    33100: contig of 884 bp in length
*    33101 33200: gap of      100 bp
*    33201    34070: contig of 870 bp in length
*    34071 34170: gap of      100 bp
*    34171    35169: contig of 999 bp in length
*    35170 35269: gap of      100 bp
*    35270    36164: contig of 895 bp in length
*    36165 36264: gap of      100 bp
*    36265    37135: contig of 871 bp in length
*    37136 37235: gap of      100 bp
*    37236    38163: contig of 928 bp in length
*    38164 38263: gap of      100 bp
*    38264    39164: contig of 901 bp in length
*    39165 39264: gap of      100 bp
*    39265    40131: contig of 867 bp in length
*    40132 40231: gap of      100 bp
*    40232    41124: contig of 893 bp in length
*    41125 41224: gap of      100 bp
*    41225    42122: contig of 898 bp in length
*    42123 42222: gap of      100 bp
*    42223    43129: contig of 907 bp in length
*    43130 43229: gap of      100 bp
*    43230    44109: contig of 880 bp in length
*    44110 44209: gap of      100 bp
*    44210    45090: contig of 881 bp in length
*    45091 45190: gap of      100 bp
*    45191    46090: contig of 900 bp in length
*    46091 46190: gap of      100 bp
*    46191    47072: contig of 882 bp in length
*    47073 47172: gap of      100 bp
*    47173    48053: contig of 881 bp in length
*    48054 48153: gap of      100 bp
*    48154    49033: contig of 880 bp in length
*    49034 49133: gap of      100 bp
*    49134    50009: contig of 876 bp in length
*    50010 50109: gap of      100 bp
*    50110    50981: contig of 872 bp in length
*    50982 51081: gap of      100 bp
*    51082    51957: contig of 876 bp in length
*    51958 52057: gap of      100 bp
*    52058    52945: contig of 888 bp in length
*    52946 53045: gap of      100 bp
*    53046    53932: contig of 887 bp in length.
FEATURES             Location/Qualifiers
     source          1. .53932
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-21D18"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    14211 a   9672 c   9208 g  15109 t   5732 others
ORIGIN


  Query Match           23.8%;  Score 77.8;  DB 2;  Length 53932;
  Best Local Similarity   55.1%;  Pred. No. 0.55;
  Matches  173;  Conservative    0;  Mismatches  138;  Indels    3;  Gaps    1;

Qy      14 taaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatat---taa 70
           |||||| |  | | ||  |  |  || |  ||| ||  |||  | || || |||   |||
Db   42654 TAAATTAATAATATAATTAATTAAAATTATAATTAATTTAAATATAATATTTATATATAA 42713

Qy      71 ttatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgttct 130
            || | |  |  |||||| ||| || ||  ||| | || ||||||||||||||  |  | |
Db   42714 ATAAAAATATATAAATTTTTAAATAAATATTAAATTTATTTAATAAATAAAATTTAATTT 42773

Qy     131 cgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaa 190
                 | ||   | || | | ||  | |||||| | |  |||   || || | ||
Db   42774 TAAATTTTATTTAAATATTATTATAAAAATAATATAATTTATTTTTTATTATAAATTTAT 42833

Qy     191 aataattatttttatttaaatttgttaaaaatttgatataattgaattgtaaaaaaaat 250
```

```
Db  42834 TATAAATATTATTATAATAATAATAAAATAATATTAAAAAATTAATATTTTTAAAATTGA 42893

Qy    251 ttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgtttt 310
Db  42894 GAAGAAATAGATTTTATATTTAATTAATTTTTATTAATTATATTAATTTTTATATAAATT 42953

Qy    311 catgttttcttatt 324
Db  42954 NAAAGATTTTAAAT 42967


RESULT   8
AC006280
LOCUS       AC006280   163443 bp    DNA             HTG       12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 1 ordered pieces.
ACCESSION   AC006280
VERSION     AC006280.8  GI:9797728
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 163443)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163443)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810453.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   163443: contig of 163443 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163443
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC724"
                     /clone="3D7"
BASE COUNT    65318 a  15421 c  17069 g  65635 t
ORIGIN


  Query Match           23.7%;  Score 77.4;  DB 2;  Length 163443;
  Best Local Similarity   52.3%;  Pred. No. 0.51;
  Matches  171;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy      1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 153862 ATTTGAGAGAATTTATATTGAAGCCCCGTCCGCTCGGTAAGGAAAAAAAAAAAAAAATAT 153921

Qy     61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 153922 AATAAATAAAATAAATAAAAAAAAAAAAAAAAATTTTATTAATAAAATAAATAAAAAAAAA 153981

Qy    121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 153982 AAAAAAATTTATTAATTAAAATTAATAAAAAAAAAAAAAAAAAAAAAATATAATAAATAA 154041

Qy    181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
```

```
Db 154042 AATAAATAAAAAAAAAAAAAAATATAATAAATAAAATAAATAAAAAATGAAATAAATAAA 154101

Qy    241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 154102 TAAAAATATATATATATATATATATATATATATTTATATATATGGGAATATATTTATCCA 154161

Qy    301 tttatgttttcatgttttcttattgtt 327
Db 154162 TTTATGTGTGGTGTTTTTCCTTTTATT 154188


RESULT   9
AC005506
LOCUS       AC005506   205429 bp    DNA             HTG       12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.
ACCESSION   AC005506
VERSION     AC005506.8  GI:9797717
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 205429)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205429)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810449.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   132269: contig of 132269 bp in length
            *  132270   132469: gap of unknown length
            *  132470   192227: contig of 59758 bp in length
            *  192228   192427: gap of unknown length
            *  192428   203864: contig of 11437 bp in length
            *  203865   204064: gap of unknown length
            *  204065   205429: contig of 1365 bp in length.
FEATURES             Location/Qualifiers
     source          1. .205429
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC357"
                     /clone="3D7"
BASE COUNT    83932 a  17937 c  19766 g  83194 t    600 others
ORIGIN


  Query Match           23.7%;  Score 77.4;  DB 2;  Length 205429;
  Best Local Similarity   52.3%;  Pred. No. 0.49;
  Matches  171;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy      1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 133530 ATTTGAGAGAATTTATATTGAAGCCCCGTCCGCTCGGTAAGGAAAAAAAAAAAAAAATAT 133589

Qy     61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 133590 AATAAATAAAATAAATAAAAAAAAAAAAAAAAATTTTATTAATAAAATAAATAAAAAAAAA 133649
```

```
Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           |||     | |         |  ||  |     |||||  |||| | ||||| |||  | |
Db  133650 AAAAAAATTTATTAATTAAAATTAATAAAAAAAAAAAAAAAAAAAAAATATAATAAATAA 133709

Qy     181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
           || |    |||| ||  |   |||  |||||    |||| |    ||  ||| |
Db  133710 AATAAATAAAAAAAAAAAAAAATATAATAAATAAAATAAATAAAAAATGAAATAAATAAA 133769

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
           |||||| |  |  |       |||||  ||  |   |  |||  |   | | || |
Db  133770 TAAAAATATATATATATATATATATATATATTTATATATATGGGAATATATTTATCCA 133829

Qy     301 tttatgttttcatgttttcttattgtt 327
           ||||||| |     |||||  | || ||
Db  133830 TTTATGTGTGGTGTTTTTCCTTTTATT 133856


RESULT  10
AF300334/c
LOCUS       AF300334     3392 bp    DNA             INV       12-NOV-2000
DEFINITION  Dictyostelium discoideum putative transcription factor StkA (stkA)
            gene, partial cds.
ACCESSION   AF300334
VERSION     AF300334.1  GI:11139097
KEYWORDS    .
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 3392)
  AUTHORS   Ke,C.H., Chen,Y.H., Liu,C.I., Yang,C.S., Gross,J.D., Newell,P.C.
            and Chang,W.T.
  TITLE     The regulation of cell fate gene stalky expression in Dictyostelium
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3392)
  AUTHORS   Ke,C.H., Yang,C.S., Huang,Y.C. and Chang,W.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2000) Biochemistry, National Cheng Kung
            University, 1 University Road, Tainan, Taiwan 70101, Republic of
            China
FEATURES             Location/Qualifiers
     source          1. .3392
                     /organism="Dictyostelium discoideum"
                     /strain="Ax2"
                     /db_xref="taxon:44689"
                     /note="44689"
     gene            1. .>3392
                     /gene="stkA"
     promoter        1. .2897
                     /gene="stkA"
     mRNA            2898. .>3392
                     /gene="stkA"
                     /product="putative transcription factor StkA"
     CDS             3390. .>3392
                     /gene="stkA"
                     /note="knock-out prevents ability to complete spore
                     formation, leading to the formation of a fruiting body
                     containing only stalk cells"
                     /codon_start=1
                     /product="putative transcription factor StkA"
                     /protein_id="AAG31607.1"
                     /db_xref="GI:11139098"
                     /translation="M"
BASE COUNT     1516 a    399 c    175 g   1302 t
ORIGIN


  Query Match             23.3%;  Score 76.2;  DB 3;  Length 3392;
  Best Local Similarity   54.6%;  Pred. No. 1.3;
  Matches  173;  Conservative    0;  Mismatches  143;  Indels    1;  Gaps    1;

Qy      11 tcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatattaa 70
           | || |  |||       | | | ||| ||| || |||| | |||||||| | | |   |
Db    1875 TTTTGATATTACTTTGATATATTTTTTTAATTAATTAAAAAAAATAAAAAAAAAAACATA 1816

Qy      71 ttatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgttct 130
           ||| |    ||||   | || | ||| | ||    | | ||||||   |||
Db    1815 ATATTTTTTGAAAAAAAAAAAAATTTATTAAAATGGAAATAAATAAAATTAAAAATACAA 1756

Qy     131 cgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaa 190
           |      | | |   | ||||    |   |  ||   | |||     |    ||
Db    1755 ATTTGATGCTCTGTGTAAATGAAAATAATATGATTTTATTTTAATTATTGTTATTATTAT 1696

Qy     191 aataattatttttatttttaaatttgttaaaaatttgatataattgaattgtaaaaaaaat 250
           || |||||| || |||||||||| || |||||||  | ||  | |||  ||||| ||
Db    1695 TATTATTATTATTTTTTTAAATTTTTTCAAAATTT-TTTTATATAAATAATAAAATAATA 1637

Qy     251 ttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgtttt 310
           || |  |    || || || ||  ||||| ||| |   ||   | |  ||||| | | ||
Db    1636 TTTACTGTTTTATCTATATTAATTAAATTGTTTTATTATTATTTTTTTTTTTTTTTTATT 1577

Qy     311 catgttttcttattgtt 327
           | |||| |||||  ||
Db    1576 TTTTTTTTTTTATTTTT 1560


RESULT  11
AC079314/c
LOCUS       AC079314   242513 bp    DNA             HTG       09-AUG-2001
DEFINITION  Homo sapiens chromosome Xp clone RP11-386I20, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC079314
VERSION     AC079314.25  GI:15042762
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 242513)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
```

```
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242513)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 31, 2001 this sequence version replaced gi:14971136.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCBA
                Center clone name: RP11-386I20
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 4% of reads
                Chemistry: Dye-terminator Big Dye: 96% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 247540 bases at least Q40
                Consensus quality: 264763 bases at least Q30
                Consensus quality: 273200 bases at least Q20
                Estimated insert size: 327546; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    65488: contig of 65488 bp in length
            *    65489    65588: gap of unknown length
            *    65589   129085: contig of 63497 bp in length
            *   129086   129185: gap of unknown length
            *   129186   161617: contig of 32432 bp in length
            *   161618   161717: gap of unknown length
            *   161718   171830: contig of 10113 bp in length
            *   171831   171930: gap of unknown length
            *   171931   175095: contig of 3165 bp in length
            *   175096   175195: gap of unknown length
            *   175196   179205: contig of 4010 bp in length
            *   179206   179305: gap of unknown length
            *   179306   184288: contig of 4983 bp in length
            *   184289   184388: gap of unknown length
            *   184389   188283: contig of 3895 bp in length
            *   188284   188383: gap of unknown length
            *   188384   192040: contig of 3657 bp in length
            *   192041   192140: gap of unknown length
            *   192141   195141: contig of 3001 bp in length
            *   195142   195241: gap of unknown length
            *   195242   198355: contig of 3114 bp in length
            *   198356   198455: gap of unknown length
            *   198456   201722: contig of 3267 bp in length
            *   201723   201822: gap of unknown length
            *   201823   205001: contig of 3179 bp in length
            *   205002   205101: gap of unknown length
            *   205102   207118: contig of 2017 bp in length
            *   207119   207218: gap of unknown length
            *   207219   209858: contig of 2640 bp in length
            *   209859   209958: gap of unknown length
            *   209959   212786: contig of 2828 bp in length
            *   212787   212886: gap of unknown length
            *   212887   215444: contig of 2558 bp in length
            *   215445   215544: gap of unknown length
            *   215545   217579: contig of 2035 bp in length
            *   217580   217679: gap of unknown length
            *   217680   220431: contig of 2752 bp in length
            *   220432   220531: gap of unknown length
            *   220532   222696: contig of 2165 bp in length
            *   222697   222796: gap of unknown length
            *   222797   225582: contig of 2786 bp in length
            *   225583   225682: gap of unknown length
            *   225683   227743: contig of 2061 bp in length
            *   227744   227843: gap of unknown length
            *   227844   230046: contig of 2203 bp in length
            *   230047   230146: gap of unknown length
            *   230147   232313: contig of 2167 bp in length
            *   232314   232413: gap of unknown length
            *   232414   234578: contig of 2165 bp in length
            *   234579   234678: gap of unknown length
            *   234679   237292: contig of 2614 bp in length
            *   237293   237392: gap of unknown length
            *   237393   239988: contig of 2596 bp in length
            *   239989   240088: gap of unknown length
            *   240089   242513: contig of 2425 bp in length.
FEATURES             Location/Qualifiers
     source          1. .242513
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Xp"
                     /clone="RP11-386I20"
BASE COUNT    71180 a  49220 c  48511 g  70846 t   2756 others
ORIGIN


  Query Match           23.3%;  Score 76.2;  DB 2;  Length 242513;
  Best Local Similarity   54.6%;  Pred. No. 0.65;
  Matches  165;  Conservative    0;  Mismatches  135;  Indels    2;  Gaps    1;

Qy      1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
          || |   || |   ||||||     | || ||| ||    | || | || |||| |   |
Db 128772 ATATAATATTTTAAAAATTTTAATAATAAAAATTTTATTTTTAATTTAATATAAAATTTA 128713

Qy     61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
            || | | ||  |  | |||| |||||| |||| |  | |||| | || |||| |
Db 128712 TTATAATAATTAATTAAATTTAAAAATTATAAATATATAAATAATATATTAAAAAAATTA 128653

Qy    121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
          | | |  |       | |      | ||||| ||      | | ||   ||| ||| |
Db 128652 ANAATAATTATTATAAATAAAANAATATTTTATAATTTNTTAAAAAATATAAAATTAAAA 128593

Qy    181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
          ||       | | || ||||| || | ||| ||  | || |||||| ||| |||  || |
Db 128592 AATNAAAATATAAAAATATTTATAATATAATTTAATAAATAATTTAATANAATAAAAATT 128533

Qy    241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
            |||| ||| |      | |   | | || ||||  |||  ||  |||| | |||||
Db 128532 ATTTAAAATTTTTATTTNTAATTTATATTTAATAAAA--ATTATAANATTTATTTTAATT 128475

Qy    301 tt 302
          ||
Db 128474 TT 128473


RESULT  12
PFMAL13P3/c
LOCUS       PFMAL13P3  318221 bp    DNA             HTG       19-AUG-1999
DEFINITION  Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL049184
VERSION     AL049184.5  GI:5763803
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
```

```
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 318221)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 24, 1999 this sequence version replaced gi:5731882.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .318221
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="13"
BASE COUNT   113610 a  29264 c  27321 g 118405 t  29621 others
ORIGIN


  Query Match             23.2%;  Score 75.8;  DB 2;  Length 318221;
  Best Local Similarity   52.0%;  Pred. No. 0.69;
  Matches  170;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
           ||||    | | ||  |   ||       ||||| |||   | |   ||| ||||  | ||
Db  141159 ATTTTTTTTTTTTTTTAAATATATACATAGAATTTTTTTTCATTAAAATATAAACATAA 141100

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
             ||| | | | |  ||||   | |  || ||| |||   |||  ||||  ||||| | |
Db  141099 AATAAAATTAATGAATAGTATATATATTGTAAATATTATATATATATATATAATATACAT 141040

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           || |          | |   | |   | || ||   |||  ||| ||| ||| || ||
Db  141039 AATGGTAGAAATATATAAATAGTATATATATATAATATAAATAATATATATAATATATAT 140980

Qy     181 aaaagtgtaaaataattatttttatttaaatttgttaaaaatttgatataattgaattg 240
           || |   ||| |||    | || |||    ||||| |  ||   |  |||||  |  |||
Db  140979 AATATATTAATATATACAATTATATGAATAATTTATATAATTATATATATATATATTTT 140920

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
             |   | || |       |  || ||| | | | | | ||  |    || || | | |||
Db  140919 ATATCTATATATATAATAGATATGTAATTTATATATAATAACTATGTGTATAATATTATT 140860

Qy     301 tttatgttttcatgttttcttattgtt 327
             ||||  ||    |  ||| |  || ||
Db  140859 ATTATTTTGAATTTCTTTTTGTTTTTT 140833


RESULT  13
AF377947/c
LOCUS       AF377947   140414 bp    DNA             HTG       15-MAY-2001
DEFINITION  Oryza sativa chromosome 3 clone OSJNBa0032E21, *** SEQUENCING IN
            PROGRESS ***, 5 unordered pieces.
ACCESSION   AF377947
VERSION     AF377947.1  GI:14040068
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 140414)
  AUTHORS   Eastman,A.P., Smith,S.C., Gingle,A., Pratt,L.H. and
            Cordonnier-Pratt,M.-M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Botany, University of Georgia, Miller Plant
            Sciences, Athens, Georgia 30602, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     6682: contig of 6682 bp in length
            *    6683     6782: gap of unknown length
            *    6783    21423: contig of 14641 bp in length
            *   21424    21523: gap of unknown length
            *   21524    48127: contig of 26604 bp in length
            *   48128    48227: gap of unknown length
            *   48228    89705: contig of 41478 bp in length
            *   89706    89805: gap of unknown length
            *   89806   140414: contig of 50609 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140414
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="3"
                     /clone="OSJNBa0032E21"
BASE COUNT    38395 a  30935 c  30737 g  39907 t    440 others
ORIGIN


  Query Match             23.1%;  Score 75.6;  DB 2;  Length 140414;
  Best Local Similarity   55.8%;  Pred. No. 0.83;
  Matches  163;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy      35 tttaaatgaaataaagataataaaaagatatattaattatatagctgaaaatttataatt 94
           || |||| || |    |||  | ||| |||||  |    || |  | || ||| |||| |
Db  21415 TTAAAATAAATTTTTAATATAATAAAAATATAAAATAATTAAAATTTAATATTAATAAAT 21356

Qy      95 atatgataagtatagttaataaataaaaagtgttctcgggggacact---tttttgtttt 151
           ||   | ||   ||      |||||||  |   | ||      | | |   ||| |  |||
Db  21355 ATTAAAAAAAATTAAAATTAAAATAAATANAATCCTAAAAAAAAAATTTATTTATAATTT 21296

Qy     152 aaaaaggaaaatataaataaaatttagataaaagtgtaaaataattatttttattttaaa 211
           | ||  ||||| || |||||||  || |||||  | ||| || |||||||  || |||  |
Db  21295 TATAAATAAAATTTATATAAAATATATATAAATATATAATATTATTATTTATAATTTTTA 21236

Qy     212 tttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaatataaatga 271
           |   ||  ||||   || |   | | || ||||||||   | |          | ||   |
Db  21235 TAAATTTAAATAATATTTNTATTATTTATAAAAAAATAATTAAATTTTTTAAAAATATA 21176

Qy     272 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttat 323
           | || || |||| |   || | |   || ||  | ||   || ||   | ||
Db  21175 ATAATATAATTTAATTATTAAATAATATATTATTTTTAAAATTTTAATTAAT 21124


RESULT  14
AC068139/c
LOCUS       AC068139   178783 bp    DNA             PRI       15-NOV-2000
DEFINITION  Homo sapiens chromosome 10 clone RP11-506M13, complete sequence.
ACCESSION   AC068139
VERSION     AC068139.5  GI:11178034
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178783)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
```

```
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178783)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 178783)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Nov 15, 2000 this sequence version replaced gi:9887668.
FEATURES             Location/Qualifiers
     source          1. .178783
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-506M13"
                     /clone_lib="RPCI-11"
BASE COUNT    53204 a  37175 c  38013 g  50391 t
ORIGIN


  Query Match             22.9%;  Score 75;  DB 9;  Length 178783;
  Best Local Similarity   53.6%;  Pred. No. 0.93;
  Matches  156;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy      14 taaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatattaatta 73
           ||||| ||  | | |  ||| | |||   | |   | |||  | | | | ||||  | ||
Db 150164 TAAATATATTATATATAAATATATAATATATAATTATATATAATATATAAATATATACTA 150105

Qy      74 atagctgaaaatttataattatatgataagtatagttaataaataaaaagtgttctcgg 133
           ||||    | | | ||||   ||||   || ||||  ||||| |||  |  | |
Db 150104 TATATAATATACTATATATACATATATAAAATATATATAATATATATTATATATAAATAT 150045

Qy     134 gggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaat 193
             | |  | | | | | | | |   |||||||| |||| || |||||| |  |  | |
Db 150044 ATTATATATGTATATATAATATATATAAATATATATAATATATAGATATATATAAATATA 149985

Qy     194 aattatttttatttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttc 253
            || || || | | || ||   |  |||||  |   | |||  |  ||||  | | | |
Db 149984 TATAATGTTAAATATATATAAATATAAAATATATAAAAATATAGATGTATTATATAATAT 149925

Qy     254 agggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttta 304
           |      |   || ||   | | || ||||   | ||   | |||| | ||
Db 149924 ATAAATTATATTATATATTATACATAATTTATGAATTATATATAATATATA 149874


RESULT  15
AC004153
LOCUS       AC004153   156060 bp    DNA             HTG       12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AC004153
VERSION     AC004153.7  GI:9797733
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 156060)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156060)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810454.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   102169: contig of 102169 bp in length
            *  102170   102369: gap of unknown length
            *  102370   156060: contig of 53691 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156060
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC812"
                     /clone="3D7"
BASE COUNT    62615 a  14889 c  15137 g  63219 t    200 others
ORIGIN


  Query Match             22.8%;  Score 74.4;  DB 2;  Length 156060;
  Best Local Similarity   54.9%;  Pred. No. 1.1;
  Matches  167;  Conservative    0;  Mismatches  136;  Indels    1;  Gaps    1;

Qy      24 acagaagaatgtttaaatgaaataaagataataaaaagatatattaattatatagctgaa 83
           | | || ||   | ||   ||  ||| | |||||||  | |||  ||  ||| |    |
Db   31682 AAATAAAAAAAATAAACAAAATAAAATAAAATAAAATAAAATAAAAAAAATAAAATAAAT 31741

Qy      84 aatttataattatatgataagtatagttaataaataaaaagtgttctcgggggacacttt 143
           ||  || ||| | || || | || | ||||||||||||| | | || |      |   | |
Db   31742 AAAATAAAATAAAATAATTA-TAAAATTAATAAATAAATATTATTATTATTTCATTTTAT 31800

Qy     144 tttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaataattatttttt 203
             |     |||||  ||||  |||||| | |||   ||    | ||  || | | ||||
Db   31801 AATTAAAAAAAAAAAAAAAAATTAAATATATTTTTTTTATTTTTATATTATTAATTTTTTG 31860

Qy     204 attttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaat 263
            |||| | ||| ||  || ||| | |  ||   |||||  | |||   | |      |||
Db   31861 TTTTTCATTTTATTGTAATTTTTTTTTTCTTTTTTTGTATAGAAATCATAATATAAAAAT 31920

Qy     264 ataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttat 323
           |  |   ||| ||  || || |   |||  | |  | | |   || |  | || | || |
Db   31921 AACATAAAAATAACATAATTTATTCTTTTATTTTTTATCTTATTTATTTTTATTATTTTTT 31980

Qy     324 tgtt 327
             ||
Db   31981 ATTT 31984


Search completed: January 24, 2002, 02:24:45
Job time: 4338 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 01:10:08 ; Search time 2272.52 Seconds
(without alignments)
1546.244 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	91	27.8	1101	13	CNS00EVL	AL069706 Drosophil
c	2	89.2	27.3	1225	13	CNS0161D	AL106171 Drosophil
c	3	85.4	26.1	1101	13	CNS003BD	AL064091 Drosophil
c	4	84.6	25.9	1101	13	CNS0039G	AL063921 Drosophil
	5	84.4	25.8	905	13	CNS00KHX	AL077798 Drosophil
c	6	83.2	25.4	524	13	CNS01U9O	AL167541 Tetraodon
	7	83	25.4	1101	13	CNS008X3	AL052544 Drosophil
	8	83	25.4	1101	13	CNS00EO7	AL069440 Drosophil
c	9	82.8	25.3	987	13	CNS014PQ	AL104456 Drosophil
c	10	82.2	25.1	1101	13	CNS001FB	AL060732 Drosophil
c	11	81.2	24.8	734	13	CNS010MP	AL099163 Drosophil
	12	81.2	24.8	1200	13	CNS016CO	AL106578 Drosophil
c	13	81	24.8	1125	10	AL547503	AL547503 AL547503
	14	80.6	24.6	1092	13	CNS020K7	AL175696 Tetraodon
c	15	80.4	24.6	973	13	CNS008OF	AL052232 Drosophil
	16	80.4	24.6	1001	13	CNS0155H	AL105023 Drosophil
	17	80.4	24.6	1101	13	CNS00FMC	AL070972 Drosophil
	18	80.2	24.5	836	13	CNS02WQ2	AL217379 Tetraodon
c	19	79.6	24.3	694	13	AQ853360	AQ853360 nbxb0003G
c	20	79.4	24.3	1101	13	CNS00EVL	AL069706 Drosophil
	21	79	24.2	966	13	CNS005ZC	AL061991 Drosophil
c	22	79	24.2	1101	13	CNS0021J	AL061936 Drosophil
c	23	78.6	24.0	581	13	CNS034DK	AL227297 Tetraodon
	24	78.6	24.0	928	13	CNS00DKY	AL071865 Drosophil
	25	78.6	24.0	996	13	CNS00FUH	AL071063 Drosophil
	26	78.4	24.0	1069	13	CNS0107G	AL098614 Drosophil
c	27	78.4	24.0	1101	13	CNS00EO7	AL069440 Drosophil
	28	78.4	24.0	1101	13	CNS016LI	AL106896 Drosophil
c	29	78	23.9	876	13	CNS009G1	AL053529 Drosophil
	30	77.8	23.8	987	13	CNS014PQ	AL104456 Drosophil
c	31	77.6	23.7	1101	13	CNS00FYG	AL071206 Drosophil
c	32	77.6	23.7	1101	13	CNS00Z38	AL097166 Drosophil
c	33	77.4	23.7	770	13	AQ740708	AQ740708 HS_5507_A
c	34	77.4	23.7	1101	13	CNS000B8	AL063632 Drosophil
c	35	77.4	23.7	1101	13	CNS00EPO	AL069493 Drosophil
	36	77.4	23.7	1225	13	CNS0161D	AL106171 Drosophil
	37	77.2	23.6	1101	13	CNS00KAE	AL077628 Drosophil
c	38	77.2	23.6	1200	13	CNS016CO	AL106578 Drosophil
	39	76.8	23.5	614	13	CNS0152H	AL104915 Drosophil
c	40	76.8	23.5	1099	10	AL536986	AL536986 AL536986
c	41	76.6	23.4	1101	13	CNS0039Q	AL063931 Drosophil
c	42	76.4	23.4	1092	13	CNS020K7	AL175696 Tetraodon
c	43	76.2	23.3	928	13	CNS00DKY	AL071865 Drosophil
c	44	76.2	23.3	1203	13	CNS015WU	AL106008 Drosophil
	45	76	23.2	1124	13	CNS073BM	AL427304 clone BA0

ALIGNMENTS

RESULT 1
CNS00EVL
LOCUS CNS00EVL 1101 bp DNA GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29B23"
                     /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN
```

```
  Query Match            27.8%;  Score 91;  DB 13;  Length 1101;
  Best Local Similarity  41.6%;  Pred. No. 0.0093;
  Matches  127;  Conservative   68;  Mismatches  108;  Indels    2;  Gaps    1;

Qy      15 aaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatattaattat 74
           |||    :  : | || :|| : |||:| ||| :|| | ||:||:|| : ::|| |  }::
Db     700 AAAAAAWATWAAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATATWW 759

Qy      75 atagctgaaaatttataattatatgataagtatagttaataaataaaaagtgttctcggg 134
           |||  | :  :| :::| ::|::: |||: ||:| | :| :|: :| || | :
Db     760 ATATATAWTTWTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATAWATW 819

Qy     135 ggacacttttttgtt--ttaaaaaggaaaatataaataaaatttagataaaagtgtaaaa 192
             | |  : | :  |  |  |::| |:|: | ||||: ||| |: ||:| | :  ||||
Db     820 AAAWAWAWATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAAAAAA 879

Qy     193 taattatttttatttaaatttgttaaaaatttgatataattgaattgtaaaaaaaattt 252
           ||: : |::|: |||::|:: | | ||:|: |  | |:||   ||    || ||:|:::|
Db     880 TAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAATAAWAWWWT 939

Qy     253 caggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttca 312
             |       ||  ||:|  | ::| |:|||: |:  | || | : ||:| ||| |::| :
Db     940 WATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTATAWTTWTATWTATATATWWTAW 999

Qy     313 tgttt 317
           | : |
Db    1000 TAWAT 1004
```

```
RESULT   2
CNS0161D/c
LOCUS       CNS0161D     1225 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN15C18"
                     /note="end : SP6"
BASE COUNT      266 a    128 c     38 g    368 t    425 others
ORIGIN
```

```
  Query Match            27.3%;  Score 89.2;  DB 13;  Length 1225;
  Best Local Similarity  40.8%;  Pred. No. 0.014;
  Matches  125;  Conservative   61;  Mismatches  120;  Indels    0;  Gaps    0;

Qy       8 atatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatat 67
           |||| :  :  |||  : | |: : | ||  | :   :||:| | ||:|||| ||  ::
Db    1164 ATATAWATWTATTATWWWAAAWAWTTTTTATATWATTWTAWATAAAAWAAAAATATTAWW 1105

Qy      68 taattatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgt 127
           :|:::| : |  : :||:: |:: :|:  ||   ||    : |  ||||| | :|: : :
Db    1104 WAWWWAAWAAWAWAAWWATWWTWTWKAATATAAAARTAWATATATAAAAATWAWAWWW 1045

Qy     128 tctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtg 187
           |       | | |||||| |:||  |:|  |||| : ||:  |||||   :  ||| :
Db    1044 TWAWAAAATATATTTTTTTTTWTTTTAWAAAAAAAAAWAAAWATAAATTATWWATAAAWWW 985

Qy     188 taaaataattatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaa 247
           ||||| ||     :|||||||:|:|:|| || :|||::|  | |  || ::||  |:||:
Db     984 TAAAAAAAAATAAWTTATTTWAWAWTTTTTTRAAAWWTWTTTTTTTTTTWWTTTAAWAAWT 925

Qy     248 aatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgt 307
           ::||  : :     :|: :|:   :::   || |||     ||   | |  ||||  : |
Db     924 WWTTATWTRTTTWTRTRRRARAATWWWTTTTTTTTTTTTTTTTTATATTTTTTTTTMYHHT 865

Qy     308 tttcat 313
           :|:: |
Db     864 YTYMTT 859
```

```
RESULT   3
CNS003BD/c
LOCUS       CNS003BD     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064091
VERSION     AL064091.1  GI:4941847
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
```

```
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K08"
                     /note="end : TET3"
BASE COUNT      395 a    120 c    103 g    334 t    149 others
ORIGIN


  Query Match           26.1%;  Score 85.4;  DB 13;  Length 1101;
  Best Local Similarity   46.3%;  Pred. No. 0.043;
  Matches  151;  Conservative   40;  Mismatches  134;  Indels    1;  Gaps    1;

Qy      2 tttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaag 61
          |||    ::| ||| | |:|  | | || ||| ||: : : || :::   : | | |||
Db    998 TTTTTTTWWTATTATAATWATWATATAAAAATATTWTWTWTAATWRWTTTWTARAKAAAW 939

Qy     62 atatattaattatatagctgaaaatttataattatatgataagtatagttaataaataaa 121
          | :||:| : | | ||  |  || ||| |  |: ||:  ||: |||  : ||:: ||:|:
Db    938 AAWTAWTTWATTTTTATWT-TAATTTTTTTTTWTTAWTTTAWTTATTTWAAAWWTATWAW 880

Qy    122 aagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagata 181
          |  : || |          |||||: ||||| :   || :| ||| :|:| || |
Db    879 ATAWATTATTTTTATTATTTTTTTWTTTTTATTWTANAATWTTTAATKAWATTTAAWTAT 820

Qy    182 aaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattgt 241
          :|| | ||||: ||:|||: : || |  || |: |:| |  ||| |||||| | || :
Db    819 WAAATTTAAAWAAAWTATWNWNATATATAAATWATWATATTTTTAATATAAATAAAAWTA 760

Qy    242 aaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaattt 301
          :      |   |       | |:|:   | |   |: :|||  || |||  | | || |
Db    759 TWTTTTTTTAATATATTTWTAAAWAWTATATATTTAWAWTTTATAATTTTTTTTTTTATAT 700

Qy    302 ttatgttttcatgttttcttattgtt 327
          || | ||||  | |||| ||| | ||
Db    699 TTTTTTTTTTTTTTTTTTTAATWTT 674


RESULT   4
CNS0039G/c
LOCUS       CNS0039G     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN


  Query Match           25.9%;  Score 84.6;  DB 13;  Length 1101;
  Best Local Similarity   26.0%;  Pred. No. 0.053;
  Matches   85;  Conservative  134;  Mismatches  108;  Indels    0;  Gaps    0;

Qy      1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
          ::|:   ::|   |:::::::: : || ||: :: |: ::::::|::::|:|:: |:::
Db    765 WKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDD 706

Qy     61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
          :: : :: |:     : | ::::::  |::: : ::|: |::  |:|: :::::::: :
Db    705 RRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRK 646

Qy    121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           |:: : |      : : :  ::: : :::|:|:::::::||:|:|:|::|:::: :|:|:
Db    645 GARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTWTDARKADRDWAKARAW 586

Qy    181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
          :|:::   ||:|::::| : :||  |:| :|    ::|::|: || ||:|  :|  :||
Db    585 RARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATTTATWTTTWTTTWTTT 526

Qy    241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
          |       :||| |      :   :||:| ||:|||| |:       |||  | :|:|
Db    525 TTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTA 466

Qy    301 tttatgttttcatgttttcttattgtt 327
            ::| :|:|  | |||| : ||| ||
Db    465 AAWWTAWTWTTTTWTTTTWWAATTTTT 439


RESULT   5
CNS00KHX
LOCUS       CNS00KHX     905 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR17N06 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077798
VERSION     AL077798.1  GI:4957174
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 905)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
```

```
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .905
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR17N06"
                     /note="end : T7"
BASE COUNT      388 a      82 c      77 g     194 t     164 others
ORIGIN


  Query Match             25.8%;  Score 84.4;  DB 13;  Length 905;
  Best Local Similarity   43.3%;  Pred. No. 0.063;
  Matches  141;  Conservative   55;  Mismatches  129;  Indels    1;  Gaps    1;

Qy       2 tttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaag 61
           :|:  ::||:   ||| : |    | ':  |   :  ||  ||| | | ||:| |||||
Db     581 DTKWWKWTAWKAAAAAAHAAWATAAATWATAATAWATAAAAAAAAATAAATWAAAAAAAA 640

Qy      62 atatattaattatatagctgaaaatttataattatatgataagtatagttaataaataaa 121
           | | || ||::||| |    | : :|||:||  | |  | |:  |:|   || ||| |||
Db     641 AAAAATAAAWWATAAAAAATATWTWTTAWAAAAAAAAAAAAAWAAAWAWAAAAAAAAAAAA 700

Qy     122 aagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagata 181
           ||              | | :|| :: |::: |::|  |:|: |:||| | ||| :: |:
Db     701 AAAAAAAAATAAAAWATAWWTTAWWTTWWWTAWWAWTAWAWAAWAAAAATAATAWWAAWT 760

Qy     182 aaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattgt 241
           ||:  :: | |:|  |||:   | || :|:| | :||::|  :| |||::|   |: :
Db     761 AAW-ARWGATAWATATATWAAAAATTAWAWTATAWTAHWATAWTTATAWWAAATAWAWAA 819

Qy     242 aaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaattt 301
           ||:| |:|  : :      |: |||||: |||||||  |:   ||  | |:   |  |||
Db     820 AAWATAWAAAWWWATATAAAWAATAAAWAAAAAAAAAAAWAAAAATATATWWAATTTTTT 879

Qy     302 ttatgttttcatgttttcttattgtt 327
           || | |:|| :| |||| || || ||
Db     880 TTTTWTWTTTWTTTTTTTTTTTTTT 905


RESULT   6
CNS01U9O/c
LOCUS        CNS01U9O      524 bp    DNA             GSS       12-MAY-2000
DEFINITION   Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
             196C24 of library G from Tetraodon nigroviridis, genomic survey
             sequence.
ACCESSION    AL167541
VERSION      AL167541.1  GI:7805598
KEYWORDS     GSS; genome survey sequence.
SOURCE       Tetraodon nigroviridis.
  ORGANISM   Tetraodon nigroviridis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Tetraodon.
REFERENCE    1  (bases 1 to 524)
  AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
             Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
             Weissenbach,J.
  TITLE      Charaterization and repeat analysis of the compact genome of the
             freshwater pufferfish Tetraodon nigroviridis
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 524)
  AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
             Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
             Saurin,W. and Weissenbach,J.
  TITLE      Human gene number estimate provided by genome wide analysis using
             Tetraodon nigroviridis DNA sequence
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 524)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT      This sequence is a single read and was generated as part of a large
             scale clone-end sequencing project of the Tetraodon nigroviridis
             genome. For more information, please take a look at
             http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="196C24"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG196BB12SP1~end :
                     PUC-Ori"
BASE COUNT      124 a      15 c      24 g     298 t      63 others
ORIGIN


  Query Match             25.4%;  Score 83.2;  DB 13;  Length 524;
  Best Local Similarity   43.2%;  Pred. No. 0.12;
  Matches  133;  Conservative   46;  Mismatches  129;  Indels    0;  Gaps    0;

Qy      15 aaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatattaattat 74
           |||   |  | : |  | : : ||||| |:| |||  ||| |::|: | | :  || :||
Db     399 AAAAAAATAAAWAATWATWTWATAAATWAWAAAAAATTAAAAWWAWTAAAAWAAAAAWAT 340

Qy      75 atagctgaaaatttataattatatgataagtatagttaataaataaaagtgttctcggg 134
             |   :  : |||:||| ||  :|  |:  |||  | || |:||||: |
Db     339 TTTAAWTTWTATTWATATTTTAWTTTTWTTTATTTTAAAAAWATAAWTAWAAAAAARAAA 280

Qy     135 ggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaata 194
             | |            |||||  |||||| ||| ||||  |  | |:|:    ||: |
Db     279 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAATTAAAAWAWAAAAAAWTTT 220

Qy     195 attatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttca 254
           |||| || |||  |  || | :|: || |  |    || ||    :|||||||    |
Db     219 TATATTATTTTTTATATATTTTAWAWTATATTTTTATTTTAAAAAAAWAAAAAAAAAAAA 160

Qy     255 ggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatg 314
             ||:|||||  |||||| || ||| ::: ||:: | :::::::::: :::: :
Db     159 AAAAWTAAWATAAAAAAAAAAATTTTTTTTWWWTTTWWTTTWWWWWWWWWWWWWWWWST 100

Qy     315 ttttctta 322
           |||| :::
Db      99 TTTTWWWW 92


RESULT   7
CNS008X3
LOCUS        CNS008X3     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION   Drosophila melanogaster genome survey sequence T7 end of BAC #
             BACR18L14 of RPCI-98 library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION    AL052544
VERSION      AL052544.1  GI:4934295
KEYWORDS     GSS.
SOURCE       fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 1101)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
             please see  http://www.fruitfly.org The BDGP Drosophila
             melanogaster BAC library was prepared by Kazutoyo Osoegawa and
```

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR18L14"
                     /note="end : T7"
BASE COUNT      251 a    151 c    169 g    239 t    291 others
ORIGIN
```

```
  Query Match             25.4%;  Score 83;  DB 13;  Length 1101;
  Best Local Similarity   37.3%;  Pred. No. 0.082;
  Matches  113;  Conservative   74;  Mismatches  116;  Indels    0;  Gaps    0;

Qy      13 ttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatattaatt 72
           ::||   ||  :   ||:||::   ||: :||||| | | ||:| || | :: :: || :
Db     799 DDAAKAATAAAWAGAAADAADDAAAAARAKAAATAKAAAGAAKATAAKGKKRARKGAAGK 858

Qy      73 atatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgttctcg 132
            | |:    :||| |: ||:|| :|  ::|: :| |   :|:||| |||| :
Db     859 TTTTWAWAAWAAAATWTTAWTTTWAATWWAWADAAAAAAWADAAAAAAAAKADAAATATA 918

Qy     133 ggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaa 192
            : :| |  | :: : |  ||: |  |:|:||  | : : ||::  | ||:|  : : |
Db     919 ARAKAAATATAWWADATAAAADTATAAWAWTAATATWKWTATWWTAAAAAWAAAKARTAT 978

Qy     193 taattatttttatttaaatttgttaaaaatttgatataattgaattgtaaaaaaaattt 252
           :: |:|||||| | : ||   ::|:|||  | | | ||: : |:: | | :|  :::
Db     979 ATRWAAWTTTTATATAWGATGAAWWAWAAAAGTAAAAAAAKARTAWWWTTATWATGWWWA 1038

Qy     253 caggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttca 312
           :      || |:: || | | :::|:: :  |:|:: :: |||  |||||::||||: :
Db    1039 WWAAAAAAAAAADWTATAATATWWWTDWGWRWAWAKKAWWKTGTTGTTTTTWKGTTTKTW 1098

Qy     313 tgt 315
           | |
Db    1099 TWT 1101
```

```
RESULT   8
CNS00EO7
LOCUS       CNS00EO7     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
```

please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29P01"
                     /note="end : TET3"
BASE COUNT      366 a     66 c    104 g    351 t    214 others
ORIGIN
```

```
  Query Match             25.4%;  Score 83;  DB 13;  Length 1101;
  Best Local Similarity   45.0%;  Pred. No. 0.082;
  Matches  144;  Conservative   48;  Mismatches  119;  Indels    9;  Gaps    1;

Qy       8 atatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatat 67
           || | || | :::|    | || ||| |     | :| |  | ||| |:||:  : | |
Db     557 ATKTTTTTATWWWATATAAAAAWAATTTAATTTTTWATTTTATATATTWAAWGWWAAAAA 616

Qy      68 taattatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgt 127
            |::|||| :  : ||| ||   : |:|:||  ||| ||:| ||| | ||:  |:: : |
Db     617 AAWWTATAAWHTWTAAATTTATAWTTWAWATTTTAAATAWATTTATTTAAWTTAWWTWTT 676

Qy     128 tctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtg 187
           |          |  :|||  :| | |||:  ||  :| ||: |||:|||: ||||:|
Db     677 TA---------AWAWTTTAAWTATTAAARWWAATTWAAAAWAAAAWTTTWAATAAWAAAT 727

Qy     188 taaaataattatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaa 247
           |||||||| | |||||  |||||:: : | |||   :| ||: |: :  :|   ||||||
Db     728 AAAAATAAATTTTTTTTATTTAAWWAWTTAAAATTAWTWATWAAWAWTTWTAWAAAAAAA 787

Qy     248 aatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgt 307
           |  :|        :::|  |||  | |:: |||: |  :|  | | | : | |::| | |
Db     788 ATNWTATAAWATTWWKAATAATWTATAWWTTTAWATATWATRTAAATTWTAATWYTTTTT 847

Qy     308 tttcatgttttcttattgtt 327
           :|| |: ||    | || ||
Db     848 WTTTAWTTTAAWATTTTTTT 867
```

```
RESULT   9
CNS014PQ/c
LOCUS       CNS014PQ      987 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104456
VERSION     AL104456.1  GI:5616067
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
```

```
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES              Location/Qualifiers
     source           1. .987
                      /organism="Drosophila melanogaster"
                      /plasmid="pBeloBAC11"
                      /db_xref="taxon:7227"
                      /clone_lib="DrosBAC"
                      /clone="BACN12P22"
                      /note="end : SP6"
BASE COUNT      257 a    122 c    122 g    241 t    245 others
ORIGIN


  Query Match             25.3%;  Score 82.8;  DB 13;  Length 987;
  Best Local Similarity   37.0%;  Pred. No. 0.093;
  Matches  118;  Conservative   79;  Mismatches  121;  Indels    1;  Gaps    1;

Qy       8 atatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatat 67
           :::: :::::|||   :        | || |::|  |:|::: ::: |  ::: :|:| |
Db     952 WWWWWWWWWWTTTTTTWTTTTTTTTTTTTAAWWTTTAWTWWWWWHWTTTYHWWWWTWTTT 893

Qy      68 taattatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgt 127
           |:: |:: |   | : :: :: ::  | |:| ::||  ||   ||::||| ||||| :
Db     892 TWWATWWTTTWATTWTWWAWYCYWTATTTWTWWWAAWAATTTATAWWAAAAAAAAAWWAA 833

Qy     128 tctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtg 187
           ::|      : | |  ||: : |:|| :|  :|| ||:  |   || |:| ::|||| |
Db     832 WMTNTAWWAWTATTAATTWWWATWAATWAATWAATTAWTTAATTAA-TWAAWWAAAAATT 774

Qy     188 taaaataattatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaa 247
           |:|| | | || :||| ||| ||:: | :|||:|:||| | :: || : |  ||||||
Db     773 TWAATTTAATAAWTTTCTTTAAAWWATWWTAAWAWTTTWTTTWWTTTTWTTAAAAAAAAA 714

Qy     248 aatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgt 307
           ||: : :       | :|::|| :  |   || |||  :  |||  | |  |  || |
Db     713 AAWAWWWWAAWAWNANWTWWATAWTTATTTTTTTTTTTWTWTTTTTTATTCTAATTTTWG 654

Qy     308 tttcatgttttcttattgt 326
           |::  : |:|| |: |:::
Db     653 TWWTKWTTKTTKTKKTKKK 635


RESULT  10
CNS001FB/c
LOCUS       CNS001FB     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060732
VERSION     AL060732.1  GI:4939397
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR04A23"
                      /note="end : TET3"
BASE COUNT      288 a    110 c    103 g    491 t    109 others
ORIGIN


  Query Match             25.1%;  Score 82.2;  DB 13;  Length 1101;
  Best Local Similarity   43.4%;  Pred. No. 0.1;
  Matches  135;  Conservative   42;  Mismatches  134;  Indels    0;  Gaps    0;

Qy      17 atttagcacagaagaatgtttaaatgaaataaagataataaaaagatatattaattatat 76
           || ||  :   || |:: | ||| : | |:::| ||:|  |||| ||: | |:|| :
Db    1101 ATATATAWATAAAAAWWTTATAATWTATAWWWAAATWAATAAAATATWAAATWATAWATA 1042

Qy      77 agctgaaaatttataattatatgataagtatagttaataaataaaaagtgttctcggggg 136
           |    |||| |   |||:||   | |   |   | || |:||:|||:      |
Db    1041 ATAAAAAAAATATAAATWATTAAAAATAAAATWTAAAAAWATWAAAWAAAAAATTWAATA 982

Qy     137 acacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaataat 196
           |   :| | |     |||||  |||| | |   | ||:: | | |||| : ||| |  ||
Db     981 ATTAWTATATTNAAAAAAAAWAAAAAAAAATTAATAAWWAAAAAAAAAAAWATAATAATAT 922

Qy     197 tatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttcagg 256
           ::   |: |||:  ||:|||  | |||:||  :|   |||:|||||| :
Db     921 ATAAAWWTAATWTAATWNATTWAAAAAATAATAWAATAWWAAATTAAWAAAAATAWATWA 862

Qy     257 ggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgtt 316
           : :||:: || |:  :||  ||:|  :| :: |   |::|||||   | |  || |
Db     861 AAAAWNTAWWTATAAWTTWAAAAATWTNTWAWWWAAAAATWWTTTTTTAATATATATWTA 802

Qy     317 ttcttattgtt 327
           | | ||: ::
Db     801 ATWTAATWTWW 791


RESULT  11
CNS010MP/c
LOCUS       CNS010MP      734 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099163
VERSION     AL099163.1  GI:5610774
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
```

```
          d'Etude du Polymorphisme Humain) with funding provided by a MRC
          project grant. The DNA was prepared from embryos by Alain Bucheton
          and Genevieve Payan. It has been constructed in the vector
          pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN04L20"
                     /note="end : T7"
BASE COUNT      288 a     62 c      2 g    211 t    171 others
ORIGIN


  Query Match             24.8%;  Score 81.2;  DB 13;  Length 734;
  Best Local Similarity   39.9%;  Pred. No. 0.17;
  Matches  110;  Conservative   60;  Mismatches  106;  Indels    0;  Gaps    0;

Qy      52 taataaaaagatatattaattatatagctgaaaatttataattatatgataagtatagtt 111
           |: |  : |  ||: ::  :||| |:  |   ::||:|||:|| | |  |:  | : |::
Db     359 TWTTTTWTATTTAWTWWTTWTATTTWTTTTTTWWTTKATAWTTTTGTATTWTTTTDTGKW 300

Qy     112 aataaataaaaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataa 171
           :| |   | | :|: || :           ::||| :|:|| |::  ::  | | ' ::|
Db     299 WAAATTAATATWGWTTTTKATTTTTTGTTGWWTTTTWTKTATAWWWWWWKTTTTTKWWAT 240

Qy     172 aatttagataaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatata 231
           | |:|:  |   |   ||| |||||: :||:| :||| :|::: |:| :|:    ||:|:
Db     239 ATTKTWTTTKTTAAAWTAATATAATWTWTTWTTWTTTTWAWKWTTWATWAWAAAAATWTR 180

Qy     232 attgaattgtaaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagttt 291
           |:: ::||  ||||||  ||| |  :   || ||| :|  :|||   |  | || ' ||:
Db     179 AKKWWWTTTAAAAAAATTTTTAAATKAAAAAAATATWTTKDAAAAKAAAAATWAATATTW 120

Qy     292 actgtaatttttatgttttcatgttttcttattgtt 327
           : ||::|||::| ||||     |||   ::|:||
Db     119 TKDWTAWWTTTWWTTTTTTKKAAATTTGAAWKTKTT 84


RESULT  12
CNS016CO
LOCUS       CNS016CO     1200 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106578
VERSION     AL106578.1  GI:5622626
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN15E04"
                     /note="end : T7"
BASE COUNT      351 a     188 c     167 g     307 t     187 others
ORIGIN


  Query Match             24.8%;  Score 81.2;  DB 13;  Length 1200;
  Best Local Similarity   39.0%;  Pred. No. 0.13;
  Matches  120;  Conservative   62;  Mismatches  126;  Indels    0;  Gaps    0;

Qy       8 atatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatat 67
           |:||   ||||||   | | ||  | :  : ||| : |   | | | |  ||: ||  ||
Db     871 AWATTAAAAATTTTYTAWAAAAATAADGAWTAATTWTAACGAAAAATTTTAAWAATTAAT 930

Qy      68 taattatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgt 127
           |::: |:| :    : : ||| :|||  ||  :   |:|| ||   ||::|||| | |
Db     931 TWWWAAYAAWAAAAAWTWTTTTWAATTTAATWTWTYTTWTAATTTTAAAWWAAAAAWTTT 990

Qy     128 tctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtg 187
           | |        : :|  :   | :|||:    :|||||||| | |: |:|: |:|:|: :
Db     991 TTTTTWAWATAWAWTAAWAWATAWAAAWTATTWAATATAAAAATAWATTWWAWAWAWAWT 1050

Qy     188 taaaataattatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaa 247
           ::||: ||  |: : |: :  ||| :    :::||    | |:|::| |||   ::|:|:
Db    1051 WWAAWAAAAAAWAWATWAWAAAAAAWAAAAWWWAAAAAAAAAWAWWTTAATAWAWWAWAW 1110

Qy     248 aatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgt 307
           |: :      || |: |:  |:|  : || ::| :|: |||   | || ::|:: |
Db    1111 TAWAWATATTWTAAAAAWTAWATAWATTWTTTTWWTWWAWATTTATATTTATAWWTWWAT 1170

Qy     308 tttcatgt 315
           | : || |
Db    1171 TAWWATAT 1178


RESULT  13
AL547503/c
LOCUS       AL547503     1125 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL547503 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI011YB03 5
            prime, mRNA sequence.
ACCESSION   AL547503
VERSION     AL547503.1  GI:12881624
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1125)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .1125
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI011YB03"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
```

```
                    http://fulllength.invitrogen.com"
BASE COUNT      415 a    142 c    148 g    345 t     75 others
ORIGIN


  Query Match             24.8%;  Score 81;  DB 10;  Length 1125;
  Best Local Similarity   46.0%;  Pred. No. 0.14;
  Matches  144;  Conservative   28;  Mismatches  141;  Indels    0;  Gaps    0;

Qy      15 aaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatattaattat 74
           |||   |  | | || | | |||  :| :::|  : | || |||||| | | |  |:  ::
Db    1079 AAAAAAAAAAAAAAAAATTTTTTTTWTWWWWTTTWWAAAAAAAAAAAAAAAAAAAAWAAWW 1020

Qy      75 atagctgaaaatttataattatatgataagtatagttaataaataaaaagtgttctcggg 134
             |    || ||::||: || ||    :|  | |   || |:|  | || |
Db    1019 TAAAAAAAATATWWATWTTTTTAATTAWAAAAAAAAAAAAAAWAATATAATTWAAAAAAWA 960

Qy      135 ggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaata 194
             ||:||| |:||| ||| |||| | ||| ||||: :|  |        |||: |
Db    959 WWTATTTTWTTTTTWTTATAAAAAAAAAAAAAAAAAAAAAWAWAATTTTTTAAAAAAAWAA 900

Qy      195 attatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttca 254
            |  |  |  | |||     | :: ||||  | |:| ||    | || ||:|||||||    :
Db    899 AAAAAATAAAATTTTNTAATAWWTAAAAAATWAWAAAAAATANTTATAWAAAAAAAAAAW 840

Qy      255 ggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatg 314
             |||:||| : ||:| | |:|| : |||  ||  |     |||| | ||||
Db    839 AWWAAAAATWTAATWAAAWANATTWATAWAAAAAATTTTTNANTATTTTTTTTTTTTTNN 780

Qy      315 ttttcttattgtt 327
            | || || || ||
Db    779 TATTTTTNTTTTT 767


RESULT  14
CNS020K7
LOCUS       CNS020K7     1092 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1-end : T7"
BASE COUNT      383 a    169 c    165 g    262 t    113 others
ORIGIN


  Query Match             24.6%;  Score 80.6;  DB 13;  Length 1092;
  Best Local Similarity   41.2%;  Pred. No. 0.16;
  Matches  132;  Conservative   60;  Mismatches  125;  Indels    3;  Gaps    1;

Qy       8 atatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatat 67
           ||:| | |:::: |  | | || ||: ||| | |:| ||| |:|||::::::: :|  |
Db     619 ATWTTTAAWWWWAAAAAAAAAAAAAWWTTTTTTTTAWAAAAAAAWAAWWWWWWWTTAAA 678

Qy      68 taattatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgt 127
            ||  |  |   |     ||| |  ||| |  | ||  | |   : :|| :|||::
Db     679 AAAAAAATTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAWTWAATWAAAWWAAAA 738

Qy     128 tctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtg 187
             :      : | ||:||| ||: ::::  ||||   :|| ::: :|||  |  |: :
Db     739 AAWAAWAAWWWAWTTWTTTTTTWAWWWWWWWAAAA---WAAAWWWTWTTAWTTTTAWWWA 795

Qy     188 taaaataattatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaa 247
           |:|::| :|  |||  |:   | |:||  :  |:     | ::|| | | |: | ||||:
Db     796 TWAWWTTWTATTTTAAAWAAAANAWTTTAWTTAWTAAAAAAWWAAATAATTWTTTAAAAW 855

Qy     248 aatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatttttatgt 307
           |:||| :      :|  : ||  |||||||||||    |:| :|:: | |  ||:|:|: |
Db     856 AWTTTAWAAATTWWAATWTAAAAAAAAAAATTAAAAAAWAAWTWWTTTTTTTTWTWAWWT 915

Qy     308 tttcatgttttcttattgtt 327
           :|  || |||| || || ||
Db     916 WTAAATWTTTTTTTTTTTT 935


RESULT  15
CNS008OF/c
LOCUS       CNS008OF     973 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR17O22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL052232
VERSION     AL052232.1  GI:4933983
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
```

```
FEATURES             Location/Qualifiers
     source          1. .973
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR17O22"
                     /note="end : TET3"
BASE COUNT      287 a    144 c    131 g    311 t    100 others
ORIGIN
```

Query Match 24.6%; Score 80.4; DB 13; Length 973;
Best Local Similarity 41.4%; Pred. No. 0.18;
Matches 123; Conservative 48; Mismatches 126; Indels 0; Gaps 0;

```
Qy      31 aatgtttaaatgaaataaagataataaaaagatatattaattatatagctgaaaatttat 90
           ||:  : :| : ||: ||:  | ||| | | |:| |  | ||| |||    | : ||| |
Db     972 AAWWAWAWATWTAAWAAAWATTTATATATAAAWAAAAAATTTANATATWATANWTTTTTT 913

Qy      91 aattatatgataagtatagttaataaataaaaagtgttctcgggggacacttttttgttt 150
           ::  |||| :|   |:|:  :|| ||| |||:: : |: :      : |  | ||| |
Db     912 WWAAATATAWTTTTTWTWAAWAAAAAAAAAAAWWTWATWTWWAAWTTWAAAATATTTNTAA 853

Qy     151 taaaaaggaaaatataaataaaatttagataaaagtgtaaaataattatttttattttaa 210
            |||:|   |   | : ::|::| | | | | |  |  :||| ||: :||:|   : ||
Db     852 AAAAWATATATTAAAWTWWAWWAATAAAAAATATATAAWAAAAAAWAWTTWTATAWATAT 793

Qy     211 atttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaatataaatg 270
           |  : || |  :|||  | |  ||  :|  ||| ||:  ::   :|    :  ||||
Db     792 AAAWATTTATTWTTTTTTTTTTTTTTWTAAAAAANAAWNAWDGTTDGTWTTWATTAAAAA 733

Qy     271 aaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 327
           |||||||   ||  ||| |||  | |  ||||||  |  :   :| ::| :: :| ::
Db     732 AAAAAAAANTTTATAAAATTTTATWTTTTTTTTANGNNKGKKKGADKTTKKKKTTKK 676
```

Search completed: January 24, 2002, 01:51:28
Job time: 2480 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 01:50:52 ; Search time 216.42 Seconds
(without alignments)
1295.376 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	327	100.0	327	20	AAX60300	Promoter of the be
	2	327	100.0	1392	20	AAX60299	DNA encoding the b
	3	130.2	39.8	936	22	AAF58252	Oligonucleotide D1
	4	130.2	39.8	936	22	AAF58254	Oligonucleotide D1
	5	130.2	39.8	936	22	AAF58257	Oligonucleotide D1
	6	130.2	39.8	936	22	AAF58259	Oligonucleotide D2
	7	130.2	39.8	936	22	AAF58262	Oligonucleotide D2
	8	130.2	39.8	938	22	AAF58255	Oligonucleotide D1
c	9	129.8	39.7	936	22	AAF58252	Oligonucleotide D1
c	10	129.8	39.7	936	22	AAF58254	Oligonucleotide D1
c	11	129.8	39.7	936	22	AAF58257	Oligonucleotide D1
c	12	129.8	39.7	936	22	AAF58259	Oligonucleotide D2
c	13	129.8	39.7	936	22	AAF58262	Oligonucleotide D2
c	14	129.8	39.7	938	22	AAF58255	Oligonucleotide D1
	15	91.6	28.0	244	22	AAF58238	Oligonucleotide D1
c	16	90.8	27.8	244	22	AAF58238	Oligonucleotide D1
c	17	66.8	20.4	4590	7	AAN60472	Sequence encoding
	18	66.4	20.3	2486	21	AAA97037	Nucleotide sequenc
c	19	64.8	19.8	20674	21	AAC58017	Arachidonic acid m
	20	63.8	19.5	2435	21	AAZ37865	Beta glucosidase (
	21	63.2	19.3	9789	17	AAT41852	cDNA encoding Plas
	22	62.2	19.0	6243	20	AAZ09598	Clostridium specie
	23	61.6	18.8	19124	18	AAT72882	Plasmodium var-7 g
	24	61.6	18.8	19124	21	AAZ98287	Plasmodium var-7 p
	25	61	18.7	700	22	AAH93026	Human inflammatory
c	26	61	18.7	5852	12	AAQ11710	Dictyostelium plas
c	27	60.4	18.5	700	22	AAH93026	Human inflammatory
c	28	60.2	18.4	1341	20	AAX61756	B. burgdorferi ant
c	29	60.2	18.4	1410	20	AAX61755	B. burgdorferi ant
	30	59.6	18.2	513445	22	AAI61373	Soybean 318013 reg
c	31	59.4	18.2	6265	20	AAX08523	NBP46 (root lectin
	32	59	18.0	381	22	AAH93355	Human chromosome 1
c	33	58.8	18.0	53585	20	AAX20251	Borrelia burgdorfe
	34	58.4	17.9	3975	9	AAN81157	Malaria-specific g
	35	58.4	17.9	3975	13	AAQ22999	SERP gene. Plasmo
c	36	58.4	17.9	163319	21	AAF22306	Arabidopsis thalia
c	37	58.2	17.8	1132	21	AAF12929	Aspergillus oryzae
	38	58.2	17.8	2503	15	AAQ53480	pNPX30 xylanase cD
c	39	58.2	17.8	9048	18	AAT43225	Brassica napus FCA
	40	57.8	17.7	8310	20	AAZ29911	cDNA encoding a SC
c	41	57.6	17.6	1907	20	AAX20307	Borrelia burgdorfe
	42	57.4	17.6	366	22	AAH93356	Human chromosome 1
	43	57	17.4	20674	21	AAC58017	Arachidonic acid m
c	44	56.8	17.4	605	17	AAT31530	Human 3' apolipopr
c	45	56.8	17.4	665	21	AAZ31996	Human apolipoprote

ALIGNMENTS

RESULT 1
AAX60300
ID AAX60300 standard; DNA; 327 BP.
XX
AC AAX60300;
XX
DT 12-AUG-1999 (first entry)
XX
DE Promoter of the beta-2 toxin gene of Clostridium perfringens type C.
XX
KW Beta-2 toxin; Clostridium perfringens type C; gene promoter;
KW vaccine; Clostridium tetani; ss.
XX
OS Clostridium perfringens.
XX
PN FR2768747-A1.
XX
PD 26-MAR-1999.
XX
PF 19-SEP-1997; 97FR-0011710.
XX
PR 19-SEP-1997; 97FR-0011710.
XX
PA (INSP) INST PASTEUR.
XX
PI Gibert M, Popoff MR;
XX
DR WPI; 1999-217498/19.
XX
PT Clostridium beta2 toxin gene promoter and signal sequence - useful
PT against toxins from Clostridium perfringens
XX
PS Claim 1; Page 32; 46pp; French.
XX

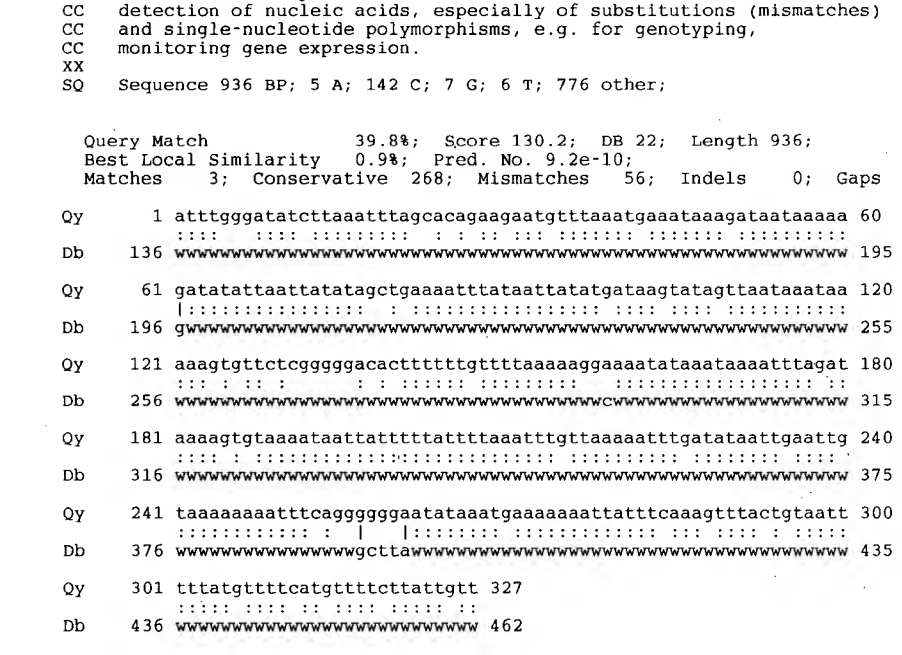
```
CC   The present sequence represents the promoter of the beta-2 toxin
CC   gene of Clostridium perfringens type C. The beta2-toxin promoter
CC   and gene sequences can be used to produce vaccines against Clostridium,
CC   and especially Clostridium perfringens, or Clostridium
CC   tetani.
XX
SQ   Sequence 327 BP; 141 A; 13 C; 44 G; 129 T; 0 other;


  Query Match             100.0%;  Score 327;  DB 20;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-35;
  Matches  327;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180

Qy     181 aaaagtgtaaaataattatttttatttaaatttgttaaaaatttgatataattgaattg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aaaagtgtaaaataattatttttatttaaatttgttaaaaatttgatataattgaattg 240

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300

Qy     301 tttatgttttcatgttttcttattgtt 327
           |||||||||||||||||||||||||||
Db     301 tttatgttttcatgttttcttattgtt 327


RESULT   2
AAX60299
ID   AAX60299 standard; DNA; 1392 BP.
XX
AC   AAX60299;
XX
DT   12-AUG-1999  (first entry)
XX
DE   DNA encoding the beta-2 toxin of Clostridium perfringens type C.
XX
KW   Beta-2 toxin; Clostridium perfringens type C; gene promoter;
KW   vaccine; Clostridium tetani; ss.
XX
OS   Clostridium perfringens.
XX
PN   FR2768747-A1.
XX
PD   26-MAR-1999.
XX
PF   19-SEP-1997;   97FR-0011710.
XX
PR   19-SEP-1997;   97FR-0011710.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gibert M,  Popoff MR;
XX
DR   WPI; 1999-217498/19.
DR   P-PSDB; AAY16591.
XX
PT   Clostridium beta2 toxin gene promoter and signal sequence - useful
PT   against toxins from Clostridium perfringens
XX
PS   Example A; Page 31; 46pp; French.
XX
```

```
CC   The present sequence encodes the beta-2 toxin of Clostridium
CC   perfringens type C. The specification describes the Clostridium
CC   perfringens beta 2 toxin gene promoter (see AAX60300). The
CC   sequences can be used to produce vaccines against Clostridium,
CC   and especially Clostridium perfringens, or Clostridium
CC   tetani.
XX
SQ   Sequence 1392 BP; 606 A; 115 C; 209 G; 462 T; 0 other;


  Query Match             100.0%;  Score 327;  DB 20;  Length 1392;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-35;
  Matches  327;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180

Qy     181 aaaagtgtaaaataattatttttatttaaatttgttaaaaatttgatataattgaattg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aaaagtgtaaaataattatttttatttaaatttgttaaaaatttgatataattgaattg 240

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300

Qy     301 tttatgttttcatgttttcttattgtt 327
           |||||||||||||||||||||||||||
Db     301 tttatgttttcatgttttcttattgtt 327


RESULT   3
AAF58252
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
```

```
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;


  Query Match             39.8%;  Score 130.2;  DB 22;  Length 936;
  Best Local Similarity   0.9%;  Pred. No. 9.2e-10;
  Matches    3;  Conservative  268;  Mismatches   56;  Indels    0;  Gaps    0;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
           ::::   :::: ::::::::::  : : :: ::: ::::::: ::::::: ::::::::::
Db     136 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 195

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
           |:::::::::::::::::  : :::::::::::::::::: :::: :::: ::::::::::
Db     196 gwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 255

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           ::: : :: :      : : :::::: ::::::::   :::::::::::::::::: ::
Db     256 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 315

Qy     181 aaaagtgtaaaataattatttttatttttaaatttgttaaaaatttgatataattgaattg 240
           :::: : :::::::::::::::::::::::::::: ::::::::: :::::::: ::::
Db     316 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 375

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
           :::::::::::: :  |   |:::::::: :::::::::::::: ::: :::: : :::::
Db     376 wwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 435

Qy     301 tttatgttttcatgttttcttattgtt 327
           ::::: :::: :: :::: ::::: ::
Db     436 wwwwwwwwwwwwwwwwwwwwwwwwwwww 462


RESULT   4
AAF58254
ID   AAF58254 standard; DNA; 936 BP.
XX
AC   AAF58254;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1875.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;


  Query Match             39.8%;  Score 130.2;  DB 22;  Length 936;
  Best Local Similarity   0.9%;  Pred. No. 9.2e-10;
  Matches    3;  Conservative  268;  Mismatches   56;  Indels    0;  Gaps    0;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
           ::::   :::: ::::::::::  : : :: ::: ::::::: ::::::: ::::::::::
Db     136 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 195

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
           |:::::::::::::::::  : :::::::::::::::::: :::: :::: ::::::::::
Db     196 gwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 255

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           ::: : :: :      : : :::::: ::::::::   :::::::::::::::::: ::
Db     256 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 315

Qy     181 aaaagtgtaaaataattatttttatttttaaatttgttaaaaatttgatataattgaattg 240
           :::: : :::::::::::::::::::::::::::: ::::::::: :::::::: ::::
Db     316 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 375

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
           :::::::::::: :  |   |:::::::: :::::::::::::: ::: :::: : :::::
Db     376 wwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 435

Qy     301 tttatgttttcatgttttcttattgtt 327
           ::::: :::: :: :::: ::::: ::
Db     436 wwwwwwwwwwwwwwwwwwwwwwwwwwww 462


RESULT   5
AAF58257
ID   AAF58257 standard; DNA; 936 BP.
XX
AC   AAF58257;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1954.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
```

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 39.8%; Score 130.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 9.2e-10;
Matches 3; Conservative 268; Mismatches 56; Indels 0; Gaps 0;



RESULT 6
AAF58259
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX

DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 39.8%; Score 130.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 9.2e-10;
Matches 3; Conservative 268; Mismatches 56; Indels 0; Gaps 0;

RESULT 7
AAF58262
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX

PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 39.8%; Score 130.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 9.2e-10;
Matches 3; Conservative 268; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 136 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 195

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 196 gwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 255

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 256 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 315

Qy 181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
Db 316 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 375

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 376 wwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 435

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 436 wwwwwwwwwwwwwwwwwwwwwwwww 462

RESULT 8
AAF58255
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 39.8%; Score 130.2; DB 22; Length 938;
Best Local Similarity 0.9%; Pred. No. 9.2e-10;
Matches 3; Conservative 268; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 136 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 195

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 196 gwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 255

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 256 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 315

Qy 181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
Db 316 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 375

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 376 wwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 435

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 436 wwwwwwwwwwwwwwwwwwwwwwwww 462

RESULT 9
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 39.7%; Score 129.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 1e-09;
Matches 2; Conservative 269; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 750 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 691

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 690 GWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 631

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 630 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWW 571

Qy 181 aaaagtgtaaaataattatttttatttttaaatttgttaaaaatttgatataattgaattg 240
Db 570 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 511

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 510 WWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 451

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 450 WWWWWWWWWWWWWWWWWWWWWWWWW 424

RESULT 10
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.

XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 39.7%; Score 129.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 1e-09;
Matches 2; Conservative 269; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 750 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 691

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 690 GWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 631

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 630 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWW 571

Qy 181 aaaagtgtaaaataattatttttatttttaaatttgttaaaaatttgatataattgaattg 240
Db 570 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 511

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 510 WWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 451

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 450 WWWWWWWWWWWWWWWWWWWWWWWWW 424

RESULT 11
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.

XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 39.7%; Score 129.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 1e-09;
Matches 2; Conservative 269; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 750 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 691

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 690 GWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 631

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 630 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWW 571

Qy 181 aaaagtgtaaaataattatttttatttttaaatttgttaaaaatttgatataattgaattg 240
Db 570 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 511

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 510 WWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 451

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 450 WWWWWWWWWWWWWWWWWWWWWWWWW 424

RESULT 12
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.

XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 39.7%; Score 129.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 1e-09;
Matches 2; Conservative 269; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 750 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 691

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 690 GWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 631

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 630 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWW 571

Qy 181 aaaagtgtaaaataattatttttatttttaaatttgttaaaaatttgatataattgaattg 240
Db 570 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 511

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 510 WWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 451

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 450 WWWWWWWWWWWWWWWWWWWWWWWWW 424

RESULT 13
AAF58262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.

XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 39.7%; Score 129.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 1e-09;
Matches 2; Conservative 269; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 750 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 691

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 690 GWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 631

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 630 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWW 571

Qy 181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
Db 570 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 511

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 510 WWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 451

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 450 WWWWWWWWWWWWWWWWWWWWWWWWW 424

RESULT 14
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.

XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 39.7%; Score 129.8; DB 22; Length 938;
Best Local Similarity 0.6%; Pred. No. 1e-09;
Matches 2; Conservative 269; Mismatches 56; Indels 0; Gaps 0;

Qy 1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db 750 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 691

Qy 61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db 690 GWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 631

Qy 121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db 630 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWW 571

Qy 181 aaaagtgtaaaataattatttttattttaaatttgttaaaaatttgatataattgaattg 240
Db 570 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 511

Qy 241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
Db 510 WWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 451

Qy 301 tttatgttttcatgttttcttattgtt 327
Db 450 WWWWWWWWWWWWWWWWWWWWWWWWW 424

RESULT 15
AAF58238
ID AAF58238 standard; DNA; 244 BP.
XX
AC AAF58238;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1250:D1102.
XX

KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 4; Page 120; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match 28.0%; Score 91.6; DB 22; Length 244;
Best Local Similarity 4.8%; Pred. No. 0.00013;
Matches 10; Conservative 167; Mismatches 31; Indels 0; Gaps 0;

```
Qy      42 gaaataaagataataaaaagatatattaattatatagctgaaaatttataattatatgat 101
           |::::::: :::::::::: ::::::::::::::::  : ::::::::::::::::: ::
Db       7 gwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 66

Qy     102 aagtatagttaataaataaaaagtgttctcgggggacacttttttgttttaaaaaggaaa 161
           :: :::: :::::::::::::: : :: :      : :|:::::: :::::::::  :::
Db      67 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwww 126

Qy     162 atataaataaaatttagataaaagtgtaaaataattatttttatttttaaatttgttaaaa 221
           ::::::::::::::::: :::::: : :::::::::::::::::::::::::::: ::::::
Db     127 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 186

Qy     222 atttgatataattgaattgtaaaaaaaa 249
           :::: :::::::: ::|  ||||| | |
Db     187 wwwwwwwwwwwwwwwwtgttaaaagaca 214
```

Search completed: January 24, 2002, 02:22:17
Job time: 1885 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 01:51:38 ; Search time 2099.46 Seconds
(without alignments)
1673.702 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 80718

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	15	4.6	42	10	AU011968	AU011968 AU011968
	2	15	4.6	42	10	AU011969	AU011969 AU011969
	3	15	4.6	42	10	AU011971	AU011971 AU011971
	4	15	4.6	42	10	AU011973	AU011973 AU011973
	5	15	4.6	49	10	AA922891	AA922891 oi51h05.s
	6	15	4.6	50	11	C01094	C01094 HUMGS000775
	7	14	4.3	24	13	AZ781748	AZ781748 2M0021M18
c	8	14	4.3	28	11	C01204	C01204 HUMGS000790
	9	14	4.3	34	10	AU038857	AU038857 AU038857
	10	14	4.3	42	10	AW333885	AW333885 S27E6 AGS
c	11	14	4.3	42	13	AZ634761	AZ634761 1M0490C17
c	12	14	4.3	46	13	AZ459612	AZ459612 1M0264O02
c	13	14	4.3	47	13	AZ345468	AZ345468 1M0080N12
	14	14	4.3	50	10	BE043289	BE043289 hk49d05.y
	15	14	4.3	50	13	TA154F10Q	AL473287 T. brucei
c	16	13	4.0	19	13	AZ331628	AZ331628 1M0059M12
c	17	13	4.0	25	13	AZ829725	AZ829725 2M0107I24
	18	13	4.0	26	13	AZ309204	AZ309204 1M0013F07
c	19	13	4.0	26	13	AZ866662	AZ866662 2M0177A18
c	20	13	4.0	26	13	TA123B12Q	AL463522 T. brucei
	21	13	4.0	27	13	AZ784620	AZ784620 2M0027P04
	22	13	4.0	28	13	AZ452653	AZ452653 1M0252E07
c	23	13	4.0	30	13	AZ623794	AZ623794 1M0461C19
	24	13	4.0	32	11	H40874	H40874 yn97e10.s1
	25	13	4.0	32	13	AZ458690	AZ458690 1M0263E04
c	26	13	4.0	34	10	AA906810	AA906810 ok73c02.s
	27	13	4.0	34	13	AZ586746	AZ586746 1M0392K22
	28	13	4.0	34	13	AZ781725	AZ781725 2M0021I14
	29	13	4.0	35	10	AW246486	AW246486 2821545.3
	30	13	4.0	35	11	D45807	D45807 HUMGS03025
	31	13	4.0	36	13	AZ314238	AZ314238 1M0030N24
	32	13	4.0	37	10	AA913140	AA913140 oi18h06.s
	33	13	4.0	37	10	AU009123	AU009123 AU009123
c	34	13	4.0	40	10	AA916625	AA916625 om05g12.s
c	35	13	4.0	40	10	AA922076	AA922076 oh08g07.s
	36	13	4.0	41	11	C00434	C00434 HUMGS000609
c	37	13	4.0	41	13	AZ777050	AZ777050 2M0011M12
	38	13	4.0	43	13	AZ371136	AZ371136 1M0122N05
	39	13	4.0	43	13	AZ464392	AZ464392 1M0273D17
	40	13	4.0	43	13	AZ575514	AZ575514 AST-T21F0
c	41	13	4.0	45	11	D20668	D20668 HUMGS01644
	42	13	4.0	46	13	AZ459612	AZ459612 1M0264O02
c	43	13	4.0	46	13	AZ834972	AZ834972 2M0117N19
	44	13	4.0	46	13	AZ991460	AZ991460 2M0275K07
	45	13	4.0	49	10	AI630064	AI630064 ad00157 P

ALIGNMENTS

RESULT 1
AU011968
LOCUS AU011968 42 bp mRNA EST 03-AUG-1998
DEFINITION AU011968 Schizosaccharomyces pombe late log phase cDNA
Schizosaccharomyces pombe cDNA clone spc06162, mRNA sequence.
ACCESSION AU011968
VERSION AU011968.1 GI:3356877
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
REFERENCE 1 (bases 1 to 42)
AUTHORS Morimyo,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces pombe
JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
FEATURES Location/Qualifiers
source 1. .42
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc06162"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of

Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT 19 a 1 c 7 g 15 t
ORIGIN

Query Match 4.6%; Score 15; DB 10; Length 42;
Best Local Similarity 100.0%; Pred. No. 9.6e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 tttttgttttaaaaa 156
||||||||||||||||
Db 22 TTTTTGTTTTAAAAA 36

RESULT 2
AU011969
LOCUS AU011969 42 bp mRNA EST 03-AUG-1998
DEFINITION AU011969 Schizosaccharomyces pombe late log phase cDNA
Schizosaccharomyces pombe cDNA clone spc06163, mRNA sequence.
ACCESSION AU011969
VERSION AU011969.1 GI:3356878
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
REFERENCE 1 (bases 1 to 42)
AUTHORS Morimyo,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces pombe
JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
FEATURES Location/Qualifiers
source 1. .42
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc06163"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT 19 a 1 c 7 g 15 t
ORIGIN

Query Match 4.6%; Score 15; DB 10; Length 42;
Best Local Similarity 100.0%; Pred. No. 9.6e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 tttttgttttaaaaa 156
||||||||||||||||
Db 22 TTTTTGTTTTAAAAA 36

RESULT 3
AU011971
LOCUS AU011971 42 bp mRNA EST 03-AUG-1998
DEFINITION AU011971 Schizosaccharomyces pombe late log phase cDNA
Schizosaccharomyces pombe cDNA clone spc06165, mRNA sequence.
ACCESSION AU011971
VERSION AU011971.1 GI:3356880
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
REFERENCE 1 (bases 1 to 42)
AUTHORS Morimyo,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces pombe
JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
FEATURES Location/Qualifiers
source 1. .42
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc06165"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT 19 a 1 c 7 g 15 t
ORIGIN

Query Match 4.6%; Score 15; DB 10; Length 42;
Best Local Similarity 100.0%; Pred. No. 9.6e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 tttttgttttaaaaa 156
||||||||||||||||
Db 22 TTTTTGTTTTAAAAA 36

RESULT 4
AU011973
LOCUS AU011973 42 bp mRNA EST 03-AUG-1998
DEFINITION AU011973 Schizosaccharomyces pombe late log phase cDNA
Schizosaccharomyces pombe cDNA clone spc06167, mRNA sequence.
ACCESSION AU011973
VERSION AU011973.1 GI:3356882
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
REFERENCE 1 (bases 1 to 42)
AUTHORS Morimyo,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces pombe
JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
FEATURES Location/Qualifiers
source 1. .42
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc06167"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"

/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT 19 a 1 c 7 g 15 t
ORIGIN

Query Match 4.6%; Score 15; DB 10; Length 42;
Best Local Similarity 100.0%; Pred. No. 9.6e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 tttttgttttaaaaa 156
||||||||||||||||
Db 22 TTTTTGTTTTAAAAA 36

RESULT 5
AA922891
LOCUS AA922891 49 bp mRNA EST 21-APR-1998
DEFINITION oi51h05.s1 NCI_CGAP_HN3 Homo sapiens cDNA clone IMAGE:1486233 3' similar to TR:Q34191 Q34191 NADH DEHYDROGENASE SUBUNIT 4. ;, mRNA sequence.
ACCESSION AA922891
VERSION AA922891.1 GI:3070200
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 49)
AUTHORS NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Dental Research, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: John Ensley, M.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .49
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1486233"
/clone_lib="NCI_CGAP_HN3"
/tissue_type="squamous cell carcinoma from base of tongue"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: tongue; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.0 kb. 5' adaptor sequence: 5' GAATTCGGCACGAG 3' 3' adaptor sequence: 5' (GA)10ACTAGTCTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 30 a 0 c 9 g 10 t
ORIGIN

Query Match 4.6%; Score 15; DB 10; Length 49;
Best Local Similarity 100.0%; Pred. No. 9.1e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 aattgtaaaaaaaat 250
|||||||||||||||
Db 31 AATTGTAAAAAAAAT 45

RESULT 6
C01094
LOCUS C01094 50 bp mRNA EST 23-JUL-1996
DEFINITION HUMGS0007754 Human adult (K.Okubo) Homo sapiens cDNA, mRNA sequence.
ACCESSION C01094
VERSION C01094.1 GI:1433324
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 50)
AUTHORS Okubo,K.
TITLE BodyMap; human gene expression database
JOURNAL Unpublished (1995)
COMMENT Contact: Okubo,K.
Institute for Molecular and Cellular Biol
Osaka University
1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
Tel: 06-877-5111(ex.3315)
Email: kousaku@imcb.osaka-u.ac.jp
Human Gene Signature, 3'-directed cDNA sequence. We are not submitting the same cDNA sequence redundantly to DDBJ since 1993. For the abundance information of clones with this sequence in this library and as well as in other 3'-directed libraries, see ' http://www.imcb.osaka-u.ac.jp/bodymap'. The sequences of the clones represented by this GS sequences is also found there.
FEATURES Location/Qualifiers
source 1. .50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human adult (K.Okubo)"
/dev_stage="adult"
BASE COUNT 24 a 6 c 5 g 15 t
ORIGIN

Query Match 4.6%; Score 15; DB 11; Length 50;
Best Local Similarity 100.0%; Pred. No. 9e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 aaaatataaataaaa 173
|||||||||||||||
Db 17 AAAATATAAATAAAA 31

RESULT 7
AZ781748
LOCUS AZ781748 24 bp DNA GSS 16-FEB-2001
DEFINITION 2M0021M18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0021M18 F, DNA sequence.
ACCESSION AZ781748
VERSION AZ781748:1 GI:12914752
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah

```
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0021  row: M  column: 18
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 24.
FEATURES              Location/Qualifiers
     source           1. .24
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0021M18"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT        9 a      0 c      3 g     12 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 13;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    198 atttttattttaaa 211
          ||||||||||||||
Db      7 ATTTTTATTTTAAA 20


RESULT   8
C01204/c
LOCUS       C01204         28 bp    mRNA            EST       23-JUL-1996
DEFINITION  HUMGS0007904 Human adult (K.Okubo) Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   C01204
VERSION     C01204.1  GI:1433434
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Okubo,K.
  TITLE     BodyMap; human gene expression database
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Okubo,K.
            Institute for Molecular and Cellular Biol
            Osaka University
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            Tel: 06-877-5111(ex.3315)
            Email: kousaku@imcb.osaka-u.ac.jp
            Human Gene Signature, 3'-directed cDNA sequence. We are not
            submitting the same cDNA sequence redundantly to DDBJ since 1993.
            For the abundance information of clones with this sequence in this
            library and as well as in other 3'-directed libraries, see '
            http://www.imcb.osaka-u.ac.jp/bodymap'. The sequences of the clones
            represented by this GS sequences is also found there.
FEATURES              Location/Qualifiers
     source           1. .28
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="Human adult (K.Okubo)"
                      /dev_stage="adult"
BASE COUNT       11 a      3 c      3 g     10 t      1 others
ORIGIN


  Query Match             4.3%;  Score 14;  DB 11;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    198 atttttattttaaa 211
          ||||||||||||||
Db     20 ATTTTTATTTTAAA 7


RESULT   9
AU038857
LOCUS       AU038857       34 bp    mRNA            EST       29-MAR-1999
DEFINITION  AU038857 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSL566, mRNA sequence.
ACCESSION   AU038857
VERSION     AU038857.1  GI:3985610
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .34
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="SSL566"
                      /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                      /dev_stage="slug"
BASE COUNT       23 a      0 c      0 g     11 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 10;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    110 ttaataaataaaaa 123
          ||||||||||||||
Db      2 TTAATAAATAAAAA 15


RESULT  10
```

```
AW333885
LOCUS       AW333885       42 bp    mRNA            EST       31-JAN-2000
DEFINITION  S27E6 AGS-1 Pneumocystis carinii f. sp. carinii cDNA 3', mRNA
            sequence.
ACCESSION   AW333885
VERSION     AW333885.1  GI:6830242
KEYWORDS    EST.
SOURCE      Pneumocystis carinii f. sp. carinii.
  ORGANISM  Pneumocystis carinii f. sp. carinii
            Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
            Pneumocystidaceae; Pneumocystis.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman
            ,J.C., Kovacs,J. and Cushion,M.
  TITLE     Expressed sequence tags from Pneumocystis carinii
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Staben C
            School of Biological Sciences
            University of Kentucky
            101 Morgan Building, University of Kentucky, Lexington, KY
            40506-0225, USA
            Tel: 606 257 2161
            Fax: 606 257 1717
            Email: staben@pop.uky.edu.
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="Pneumocystis carinii f. sp. carinii"
                     /db_xref="taxon:38081"
                     /clone_lib="AGS-1"
                     /lab_host="E. coli"
                     /note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
                     P. carinii organisms (3x10e9) from a single rat (99-1-6,
                     sacrificed on 3/17/99)  at Cincinnati VA facilities.
                     Trizol extracted RNA.  Oligo dT priming, standard
                     conditions described by  vendor, Stratagene. Further
                     details see www.uky.edu/Project/Pneumocystis/"
BASE COUNT       19 a      0 c      6 g     17 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 10;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    164 ataaataaaattta 177
          ||||||||||||||
Db      3 ATAAATAAAATTTA 16


RESULT  11
AZ634761/c
LOCUS       AZ634761       42 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0490C17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0490C17 R, DNA sequence.
ACCESSION   AZ634761
VERSION     AZ634761.1  GI:11756951
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0490  row: C  column: 17
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 42.
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0490C17"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       22 a      5 c      4 g     11 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 13;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    141 ttttttgttttaaa 154
          ||||||||||||||
Db     34 TTTTTTGTTTTAAA 21


RESULT  12
AZ459612/c
LOCUS       AZ459612       46 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0264O02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0264O02 R, DNA sequence.
ACCESSION   AZ459612
VERSION     AZ459612.1  GI:10617737
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
```

```
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0264  row: O  column: 02
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 46.
FEATURES              Location/Qualifiers
     source           1. .46
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0264O02"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       21 a      5 c      4 g     16 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 13;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    188 taaaataattattt 201
          ||||||||||||||
Db     45 TAAAATAATTATTT 32


RESULT  13
AZ345468/c
LOCUS       AZ345468       47 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0080N12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0080N12 F, DNA sequence.
ACCESSION   AZ345468
VERSION     AZ345468.1  GI:10424705
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 47)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0080  row: N  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 47.
FEATURES              Location/Qualifiers
     source           1. .47
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0080N12"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       12 a      9 c      0 g     26 t
ORIGIN


  Query Match             4.3%;  Score 14;  DB 13;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 aataaataaaaagt 125
          ||||||||||||||
Db     26 AATAAATAAAAAGT 13


RESULT  14
BE043289
LOCUS       BE043289       50 bp    mRNA            EST       08-JUN-2000
DEFINITION  hk49d05.y1 NCI_CGAP_Ov34 Homo sapiens cDNA clone IMAGE:3000009 5',
            mRNA sequence.
ACCESSION   BE043289
VERSION     BE043289.1  GI:8360342
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
```

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40RP from Gibco.

FEATURES Location/Qualifiers
source 1. .50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3000009"
/clone_lib="NCI_CGAP_Ov34"
/sex="female"
/tissue_type="borderline ovarian carcinoma"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pAMP1; mRNA made from borderline ovarian carcinoma, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 500 bp. Primary library, non-amplified. "

BASE COUNT 30 a 3 c 11 g 6 t
ORIGIN

Query Match 4.3%; Score 14; DB 10; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 aattgtaaaaaaaa 249
||||||||||||||
Db 27 AATTGTAAAAAAAA 40

RESULT 15
TA154F10Q
LOCUS TA154F10Q 50 bp DNA GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 154f10, reverse sequence, genomic survey sequence.
ACCESSION AL473287
VERSION AL473287.1 GI:11838560
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 50)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh1@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (
4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES Location/Qualifiers
source 1. .50
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="154f10"

BASE COUNT 20 a 7 c 9 g 14 t
ORIGIN

Query Match 4.3%; Score 14; DB 13; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 268 atgaaaaaaattat 281
||||||||||||||
Db 10 ATGAAAAAAATTAT 23

RESULT 16
AZ331628/c
LOCUS AZ331628 19 bp DNA GSS 29-SEP-2000
DEFINITION 1M0059M12R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0059M12 R, DNA sequence.
ACCESSION AZ331628
VERSION AZ331628.1 GI:10394503
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 19)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0059 row: M column: 12
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 19.

FEATURES Location/Qualifiers
source 1. .19
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0059M12"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT        8 a      1 c      0 g      10 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 7.7e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     166 aaataaaatttag 178
           |||||||||||||
Db      14 AAATAAAATTTAG 2


RESULT  17
AZ829725/c
LOCUS       AZ829725       25 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0107I24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0107I24 F, DNA sequence.
ACCESSION   AZ829725
VERSION     AZ829725.1  GI:12999549
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0107  row: I  column: 24
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 25.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0107I24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      3 c      4 g      4 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 7e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     301 tttatgttttcat 313
           |||||||||||||
Db      24 TTTATGTTTTCAT 12


RESULT  18
AZ309204
LOCUS       AZ309204       26 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0013F07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0013F07 F, DNA sequence.
ACCESSION   AZ309204
VERSION     AZ309204.1  GI:10349955
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0013  row: F  column: 07
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 26.
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0013F07"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
```

chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 4 a 4 c 4 g 14 t
ORIGIN

Query Match 4.0%; Score 13; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 196 ttatttttatttt 208
||||||||||||
Db 7 TTATTTTTATTTT 19

RESULT 19
AZ866662/c
LOCUS AZ866662 26 bp DNA GSS 21-FEB-2001
DEFINITION 2M0177A18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0177A18 F, DNA sequence.
ACCESSION AZ866662
VERSION AZ866662.1 GI:13068193
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 26)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0177 row: A column: 18
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 26.
FEATURES Location/Qualifiers
source 1. .26
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0177A18"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 20 a 0 c 0 g 6 t
ORIGIN

Query Match 4.0%; Score 13; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 196 ttatttttatttt 208
||||||||||||
Db 17 TTATTTTTATTTT 5

RESULT 20
TA123B12Q/c
LOCUS TA123B12Q 26 bp DNA GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 123b12, reverse sequence, genomic survey sequence.
ACCESSION AL463522
VERSION AL463522.1 GI:11834032
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 26)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhl@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (
4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES Location/Qualifiers
source 1. .26
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="123b12"
BASE COUNT 12 a 0 c 0 g 14 t
ORIGIN

Query Match 4.0%; Score 13; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 aaaaaaattattt 283
||||||||||||
Db 20 AAAAAAATTATTT 8

RESULT 21
AZ784620
LOCUS AZ784620 27 bp DNA GSS 16-FEB-2001
DEFINITION 2M0027P04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0027P04 R, DNA sequence.

```
ACCESSION   AZ784620
VERSION     AZ784620.1  GI:12920544
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0027  row: P  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 27.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0027P04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        5 a      5 c      5 g     12 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 6.8e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    137 acacttttttgtt 149
          |||||||||||||
Db     15 ACACTTTTTTGTT 27


RESULT  22
AZ452653
LOCUS       AZ452653       28 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0252E07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0252E07 R, DNA sequence.
ACCESSION   AZ452653
VERSION     AZ452653.1  GI:10609676
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0252  row: E  column: 07
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0252E07"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      2 c      2 g     10 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 6.7e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    271 aaaaaaattattt 283
          |||||||||||||
Db     16 AAAAAAATTATTT 28


RESULT  23
AZ623794/c
LOCUS       AZ623794       30 bp    DNA             GSS       13-DEC-2000
```

```
DEFINITION  1M0461C19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0461C19 R, DNA sequence.
ACCESSION   AZ623794
VERSION     AZ623794.1  GI:11745984
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0461  row: C  column: 19
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 30.
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0461C19"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      4 c      3 g      9 t
ORIGIN
```

```
  Query Match             4.0%;  Score 13;  DB 13;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     202 ttattttaaattt 214
           |||||||||||||
Db      14 TTATTTTAAATTT 2


RESULT  24
H40874
```

```
LOCUS       H40874          32 bp    mRNA            EST       31-JUL-1995
DEFINITION  yn97e10.s1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone
            IMAGE:176394 3' similar to gb:X54150_rna1 IMMUNOGLOBULIN ALPHA FC
            RECEPTOR PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   H40874
VERSION     H40874.1  GI:916926
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 674
            High quality sequence starts: 1
            High quality sequence stops: 1
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 674   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="Homo sapiens"
                     /db_xref="GDB:3838590"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:176394"
                     /clone_lib="Soares adult brain N2b5HB55Y"
                     /sex="Male"
                     /dev_stage="55-year old"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: brain; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 53. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. The adult brain RNA was
                     provided by Dr. Donald H. Gilden. Tissue was acquired
                     17-18 hours after death which occurred in consequence of a
                     ruptured aortic aneurysm. RNA was prepared from a pool of
                     tissues representing the following areas of the brain:
                     frontal, parietal, temporal and occipital cortex from the
                     left and right hemispheres, subcortical white matter,
                     basal ganglia, thalamus, cerebellum, midbrain, pons and
                     medulla."
BASE COUNT        6 a      0 c      0 g     26 t
ORIGIN
```

```
  Query Match             4.0%;  Score 13;  DB 11;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     196 ttatttttatttt 208
           |||||||||||||
Db      15 TTATTTTTATTTT 27
```

```
RESULT  25
AZ458690
LOCUS       AZ458690       32 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0263E04F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0263E04 F, DNA sequence.
ACCESSION   AZ458690
VERSION     AZ458690.1  GI:10616815
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0263  row: E  column: 04
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 32.
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0263E04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        4 a      5 c      4 g     19 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     196 ttatttttatttt 208
           |||||||||||||
Db      12 TTATTTTTATTTT 24


RESULT  26
AA906810/c
LOCUS       AA906810       34 bp    mRNA            EST       09-APR-1998
DEFINITION  ok73c02.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1519586 3'
            similar to SW:UFD1_HUMAN Q92890 UBIQUITIN FUSION DEGRADATION
            PROTEIN 1 HOMOLOG ;, mRNA sequence.
ACCESSION   AA906810
VERSION     AA906810.1  GI:3042054
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1519586"
                     /clone_lib="NCI_CGAP_GC4"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        8 a      4 c      7 g     15 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 10;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      51 ataataaaaagat 63
           |||||||||||||
Db      21 ATAATAAAAAGAT 9


RESULT  27
AZ586746
LOCUS       AZ586746       34 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0392K22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0392K22 R, DNA sequence.
ACCESSION   AZ586746
VERSION     AZ586746.1  GI:11708936
KEYWORDS    GSS.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0392  row: K  column: 22
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 34.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0392K22"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        2 a      0 c      4 g     28 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttt 208
          |||||||||||||
Db      6 TTATTTTTATTTT 18


RESULT  28
AZ781725
LOCUS       AZ781725       34 bp    DNA             GSS       16-FEB-2001
DEFINITION  2M0021I14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0021I14 F, DNA sequence.
ACCESSION   AZ781725
VERSION     AZ781725.1  GI:12914706
KEYWORDS    GSS.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0021  row: I  column: 14
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 34.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0021I14"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       13 a      4 c      4 g     13 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    209 aaatttgttaaaa 221
          |||||||||||||
Db     10 AAATTTGTTAAAA 22


RESULT  29
AW246486
LOCUS       AW246486       35 bp    mRNA            EST       07-JAN-2000
DEFINITION  2821545.3prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821545 3',
            mRNA sequence.
ACCESSION   AW246486
VERSION     AW246486.1  GI:6589479
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: 2821545.5prime
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling
            Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
            Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
            project Clone distribution: MGC clone distribution information can
            be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality
            Scores: PHRED from University of Washington Genome Center. Vector
            Trimming: cross_match from University of Washington Genome Center
            PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley
            Drosophila Genome Project. University of Washington Genome Center:
            http://www.genome.washington.edu Low Quality Sequence: 35
            contiguous PHRED high quality bases following vector sequence. Very
            Low Quality Sequence: Trace file contained 35 contiguous distinct
            peaks following vector sequence. Polyadenylation: Based upon the
            presence of a XhoI site followed by a run of 14 or more T residues
            at the beginning of the sequence, this cDNA insert was
            polyadenylated.
            Plate: LLCM7  row: B  column: 10
            High quality sequence stop: 35.
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2821545"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       11 a      4 c      1 g     19 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 10;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      35 tttaaatgaaata 47
           |||||||||||||
Db       8 TTTAAATGAAATA 20


RESULT  30
D45807
LOCUS       D45807          35 bp    mRNA            EST       20-FEB-1995
DEFINITION  HUMGS03025 Human adult lung 3' directed MboI cDNA Homo sapiens cDNA
            3', mRNA sequence.
ACCESSION   D45807
VERSION     D45807.1  GI:662761
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   Itoh,K., Okubo,K., Yosii,J., Yokouchi,H. and Matsubara,K.
  TITLE     An expression profile of active genes in human lung
  JOURNAL   DNA Research 1, 279-287 (1994)
  MEDLINE   95236275
COMMENT     Contact: Kohichi Itoh
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka, Suita, Osaka, 565, Japan
            Tel: 06-877-5111 x3910
            Fax: 06-877-1922.
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human adult lung 3' directed MboI cDNA"
                     /note="Adult human lung, 3' directed MboI"
BASE COUNT       14 a      3 c      5 g     13 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 11;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     106 atagttaataaat 118
           |||||||||||||
Db      11 ATAGTTAATAAAT 23


RESULT  31
AZ314238
LOCUS       AZ314238        36 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0030N24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0030N24 R, DNA sequence.
ACCESSION   AZ314238
VERSION     AZ314238.1  GI:10359929
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0030  row: N  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 36.
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0030N24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 8 a 3 c 5 g 20 t
ORIGIN

Query Match 4.0%; Score 13; DB 13; Length 36;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 atgttttcttatt 324
||||||||||||||
Db 2 ATGTTTTCTTATT 14

RESULT 32
AA913140
LOCUS AA913140 37 bp mRNA EST 14-APR-1998
DEFINITION oi18h06.s1 NCI_CGAP_HN1 Homo sapiens cDNA clone IMAGE:1483067 3' similar to SW:YKI3_YEAST P36079 HYPOTHETICAL 23.7 KD PROTEIN IN MDH1-VMA5 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION AA913140
VERSION AA913140.1 GI:3052532
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 37)
AUTHORS NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Dental Research, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: John Ensley, M.D., Mary May, J. Silvio Gutkind, Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .37
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1483067"
/clone_lib="NCI_CGAP_HN1"
/tissue_type="squamous cell carcinoma"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lymph node; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. 5' adaptor sequence: 5' GAATTCGGCACGAG 3' 3' adaptor sequence: 5' (GA)10ACTAGTCTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 2 a 9 c 2 g 24 t
ORIGIN

Query Match 4.0%; Score 13; DB 10; Length 37;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 196 ttatttttatttt 208
||||||||||||||
Db 3 TTATTTTTATTTT 15

RESULT 33
AU009123
LOCUS AU009123 37 bp mRNA EST 31-JUL-1998
DEFINITION AU009123 Schizosaccharomyces pombe late log phase cDNA Schizosaccharomyces pombe cDNA clone spc04510, mRNA sequence.
ACCESSION AU009123
VERSION AU009123.1 GI:3345803
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE 1 (bases 1 to 37)
AUTHORS Morimyo,M. and Mita,K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces pombe
JOURNAL Unpublished (1998)
COMMENT Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
FEATURES Location/Qualifiers
source 1. .37
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc04510"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT 13 a 6 c 3 g 15 t
ORIGIN

Query Match 4.0%; Score 13; DB 10; Length 37;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 ttaataaataaaa 122
||||||||||||||
Db 13 TTAATAAATAAAA 25

RESULT 34
AA916625/c
LOCUS AA916625 40 bp mRNA EST 10-JUN-1998
DEFINITION om05g12.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1540198 3' similar to SW:NU5M_TRYBB P04540 NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 ;, mRNA sequence.
ACCESSION AA916625

```
VERSION     AA916625.1  GI:3056017
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 1026   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1540198"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       10 a      1 c      2 g     27 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 10;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 5.9e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    106 atagttaataaat 118
          |||||||||||||
Db     21 ATAGTTAATAAAT 9


RESULT  35
AA922076/c
LOCUS       AA922076       40 bp    mRNA            EST       21-APR-1998
DEFINITION  oh08g07.s1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1457244 3'
            similar to TR:Q34096 Q34096 MURF2 PROTEIN. ;, mRNA sequence.
ACCESSION   AA922076
VERSION     AA922076.1  GI:3069385
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1457244"
                     /clone_lib="NCI_CGAP_Co8"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     colon adenocarcinoma, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        5 a      4 c      3 g     28 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 10;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 5.9e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 aataaataaaaag 124
          |||||||||||||
Db     38 AATAAATAAAAAG 26


RESULT  36
C00434
LOCUS       C00434         41 bp    mRNA            EST       23-JUL-1996
DEFINITION  HUMGS0006099 Human adult (K.Okubo) Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   C00434
VERSION     C00434.1  GI:1432664
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Okubo,K.
  TITLE     BodyMap; human gene expression database
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Okubo,K.
            Institute for Molecular and Cellular Biol
            Osaka University
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            Tel: 06-877-5111(ex.3315)
            Email: kousaku@imcb.osaka-u.ac.jp
            Human Gene Signature, 3'-directed cDNA sequence. We are not
            submitting the same cDNA sequence redundantly to DDBJ since 1993.
            For the abundance information of clones with this sequence in this
            library and as well as in other 3'-directed libraries, see '
            http://www.imcb.osaka-u.ac.jp/bodymap'. The sequences of the clones
            represented by this GS sequences is also found there.
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human adult (K.Okubo)"
                     /dev_stage="adult"
BASE COUNT       12 a      5 c      7 g     17 t
```

```
ORIGIN

  Query Match             4.0%;  Score 13;  DB 11;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     109 gttaataaataaa 121
           |||||||||||||
Db      14 GTTAATAAATAAA 26


RESULT  37
AZ777050/c
LOCUS       AZ777050       41 bp    DNA             GSS       16-FEB-2001
DEFINITION  2M0011M12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0011M12 F, DNA sequence.
ACCESSION   AZ777050
VERSION     AZ777050.1  GI:12905260
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0011  row: M  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 41.
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0011M12"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."

BASE COUNT       9 a      3 c     12 g     17 t
ORIGIN

  Query Match             4.0%;  Score 13;  DB 13;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      57 aaaagatatatta 69
           |||||||||||||
Db      16 AAAAGATATATTA 4


RESULT  38
AZ371136
LOCUS       AZ371136       43 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0122N05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0122N05 F, DNA sequence.
ACCESSION   AZ371136
VERSION     AZ371136.1  GI:10484836
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 43)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0122  row: N  column: 05
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 43.
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0122N05"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
```

and selected for ampicillin resistance."
BASE COUNT 28 a 2 c 2 g 11 t
ORIGIN

Query Match 4.0%; Score 13; DB 13; Length 43;
Best Local Similarity 100.0%; Pred. No. 5.7e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 taataaataaaaa 123
||||||||||||||
Db 18 TAATAAATAAAAA 30

RESULT 39
AZ464392
LOCUS AZ464392 43 bp DNA GSS 04-OCT-2000
DEFINITION 1M0273D17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0273D17 R, DNA sequence.
ACCESSION AZ464392
VERSION AZ464392.1 GI:10622517
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 43)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0273 row: D column: 17
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 43.
FEATURES Location/Qualifiers
source 1. .43
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0273D17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 6 a 7 c 8 g 22 t
ORIGIN

Query Match 4.0%; Score 13; DB 13; Length 43;
Best Local Similarity 100.0%; Pred. No. 5.7e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 139 actttttgtttt 151
||||||||||||
Db 30 ACTTTTTTGTTTT 42

RESULT 40
AZ575514
LOCUS AZ575514 43 bp DNA GSS 06-DEC-2000
DEFINITION AST-T21F0045 Genetrap T47D Human Breast Carcinoma Library Homo sapiens genomic 5', DNA sequence.
ACCESSION AZ575514
VERSION AZ575514.1 GI:11561825
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 43)
AUTHORS Henkel,G., Liyanage,M., Pratt,E., Huang,D., Riley,M., Bernardino,A., Durick,K. and Pollok,B.
TITLE Exon-trap tags from a T47D GenomeScreen(TM) Library
JOURNAL Unpublished (2000)
COMMENT Contact: Greg Henkel
Gene Expression
Aurora Biosciences Corp.
11010 Torreyana Road, San Diego, CA 92121, USA
Tel: 8584048436
Fax: 8584046719
Email: henkelg@aurorabio.com
Pools of cells were isolated from a GenomeScreen(TM) library. The library of cells was generated by retroviral integration of a gene tagging element consisting of: 1) A promoterless beta-lactamase proceeded by a splice acceptor as a reporter for gene expression; 2) A promoter driving neomycin resistance followed by a splice donor to trap downstream exons. 3' RACE from neomycin gene was performed using total RNA from isolated pools. Output was shotgun cloned in pAmp-1 and used to transform DH5-alpha competent bacteria. 5' ends of reported sequences were immediately preceded by splice donor from the trapping construct.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .43
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Genetrap T47D Human Breast Carcinoma Library"
/tissue_type="Carcinoma"
/cell_type="Epithelial"
/cell_line="T47D"
/note="Organ: Breast; Vector: pAmp-1; 3' RACE of total RNA from genetrap pools; shotgun clone in pAmp-1 and used to transform DH5-alpha competent bacteria."
BASE COUNT 18 a 9 c 5 g 11 t
ORIGIN

Query Match 4.0%; Score 13; DB 13; Length 43;
Best Local Similarity 100.0%; Pred. No. 5.7e+05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 aatgaaataaaga 51
||||||||||||
Db 1 AATGAAATAAAGA 13

```
RESULT  41
D20668/c
LOCUS       D20668          45 bp    mRNA            EST       30-JUL-1996
DEFINITION  HUMGS01644 Human promyelocyte Homo sapiens cDNA clone pm2268 3',
            mRNA sequence.
ACCESSION   D20668
VERSION     D20668.1  GI:501764
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y., Yoshinari
            ,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60 before and
            after induction of differentiation. A new application of 3'directed
            cDNA sequencing
  JOURNAL   Unpublished (1993)
COMMENT     Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1 Yamada-oka,Suita,Osaka 565,Japan.
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="pm2268"
                     /clone_lib="Human promyelocyte"
                     /note="Female, adult, cell_line = HL60, cell_type =
                     promyelocyte. "
BASE COUNT       17 a      6 c      6 g     14 t      2 others
ORIGIN


  Query Match             4.0%;  Score 13;  DB 11;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttt 208
          |||||||||||||
Db     34 TTATTTTTATTTT 22


RESULT  42
AZ459612
LOCUS       AZ459612        46 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0264O02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0264O02 R, DNA sequence.
ACCESSION   AZ459612
VERSION     AZ459612.1  GI:10617737
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0264  row: O  column: 02
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 46.
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0264O02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       21 a      5 c      4 g     16 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    190 aaataattatttt 202
          |||||||||||||
Db     32 AAATAATTATTTT 44


RESULT  43
AZ834972/c
LOCUS       AZ834972        46 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0117N19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0117N19 R, DNA sequence.
ACCESSION   AZ834972
VERSION     AZ834972.1  GI:13004880
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
```

```
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0117  row: N  column: 19
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 46.
FEATURES              Location/Qualifiers
     source           1. .46
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0117N19"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       15 a      7 c      9 g      15 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      30 gaatgtttaaatg 42
           |||||||||||||
Db      20 GAATGTTTAAATG 8


RESULT  44
AZ991460
LOCUS       AZ991460       46 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0275K07R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0275K07 R, DNA sequence.
ACCESSION   AZ991460
VERSION     AZ991460.1  GI:13862687
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0275  row: K  column: 07
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 46.
FEATURES              Location/Qualifiers
     source           1. .46
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0275K07"
                      /clone_lib="Mouse 10kb plasmid UUGC2M library"
                      /sex="Female"
                      /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (female) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       23 a      10 c      5 g      8 t
ORIGIN


  Query Match             4.0%;  Score 13;  DB 13;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 taaatgaaataaa 49
           |||||||||||||
Db      32 TAAATGAAATAAA 44


RESULT  45
AI630064
LOCUS       AI630064       49 bp    mRNA            EST       08-MAR-2000
DEFINITION  ad00157 Proliferating Erythroid Cells (LCB:ad library) Homo sapiens
            cDNA clone ad00157 random, mRNA sequence.
ACCESSION   AI630064
VERSION     AI630064.1  GI:4681394
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Gubin,A.N., Njoroge,J.M., Bouffard,G.G. and Miller,J.L.
  TITLE     Gene expression in proliferating human erythroid cells
  JOURNAL   Genomics 59 (2), 168-177 (1999)
  MEDLINE   99339981
COMMENT     Contact: Jeffery L. Miller
            Laboratory of Chemical Biology
            National Institute of Diabetes and Digestive and Kidney Diseases
            Building 10, Room 9B17, National Institutes of Health, Bethesda, MD
            20892, USA
            Tel: 301 402 2373
            Fax: 301 435 5148
```

```
           Email: jm7f@nih.gov
           The 'ad' library was constructed by Alexander Gubin, Ph.D. in the
           Laboratory of Chemical Biology, NIDDK, NIH. DNA Sequencing and/or
           analyses by National Institutes of Health Intramural Sequencing
           Center (NISC). More information available at:
           http://hembase.niddk.nih.gov.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="ad00157"
                     /clone_lib="Proliferating Erythroid Cells (LCB:ad library
                     )"
                     /sex="unknown"
                     /tissue_type="blood"
                     /cell_type="Erythroid Cells"
                     /cell_line="Primary Culture of Peripheral Blood
                     Mononuclear Cells"
                     /dev_stage="Progenitor; EPO responsive CD71++++"
                     /lab_host="DH5alpha"
                     /note="Organ: blood; Vector: pCRII; Site_1: EcoRI; Site_2:
                     EcoRI; Human peripheral blood mononuclear cells were
                     cultured in the presence or absence of erythropoietin.
                     CD71++++ cells arising only in erythropoietin-supplemented
                     medium were isolated by fluorescence activated cell
                     sorting.  Those cells demonstrated an average of six
                     additional doublings in suspension culture and erythroid
                     colony formation in methylcellulose.  Suppression
                     subtractive hybridization was used to construct the ad
                     library (tester=sorted CD71++++ cells, driver=unsorted
                     cells cultured without erythropoietin)."
BASE COUNT       16 a      7 c      9 g     13 t      4 others
ORIGIN


  Query Match             4.0%;  Score 13;  DB 10;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 5.4e+05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     172 aatttagataaaa 184
           |||||||||||||
Db      24 AATTTAGATAAAA 36


Search completed: January 24, 2002, 02:57:29
Job time: 3951 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 01:13:27 ; Search time 90.4 Seconds
(without alignments)
819.229 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	63.8	19.5	2435	4	US-09-306-593-1	Sequence 1, Appli
	2	62.2	19.0	6243	2	US-09-056-075-1	Sequence 1, Appli
c	3	61.8	18.9	615	4	US-08-998-416-186	Sequence 186, App
	4	61.6	18.8	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	5	59.2	18.1	5852	1	US-07-867-106-2	Sequence 2, Appli
c	6	58.2	17.8	9048	3	US-08-973-273-4	Sequence 4, Appli
c	7	57.8	17.7	636	4	US-08-998-416-1137	Sequence 1137, Ap
c	8	56.8	17.4	665	2	US-08-883-795A-36	Sequence 36, Appl
c	9	56.4	17.2	837	4	US-08-998-416-288	Sequence 288, App
	10	55.8	17.1	2750	3	US-08-617-860B-33	Sequence 33, Appl
	11	55.6	17.0	665	2	US-08-883-795A-36	Sequence 36, Appl
c	12	54.8	16.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
c	13	54.8	16.8	8920	4	US-09-150-741-1	Sequence 1, Appli
	14	54.6	16.7	2251	4	US-08-991-677-11	Sequence 11, Appl
	15	54.4	16.6	5852	1	US-07-867-106-2	Sequence 2, Appli
c	16	53.4	16.3	658	4	US-08-998-416-595	Sequence 595, App
	17	53.2	16.3	3095	6	5231168-1	Patent No. 5231168
c	18	52.8	16.1	2110	4	US-09-419-459-1	Sequence 1, Appli
c	19	52.8	16.1	3095	6	5231168-1	Patent No. 5231168
	20	52.6	16.1	1826	4	US-09-286-691-11	Sequence 11, Appl
	21	52.6	16.1	1826	4	US-09-687-147-11	Sequence 11, Appl
	22	52.6	16.1	3305	3	US-09-068-043-1	Sequence 1, Appli
	23	52.6	16.1	51952	3	US-08-947-823-1	Sequence 1, Appli
	24	52.2	16.0	8920	2	US-08-446-855A-1	Sequence 1, Appli
	25	52.2	16.0	8920	4	US-09-150-741-1	Sequence 1, Appli
c	26	52	15.9	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	27	51.2	15.7	1186	2	US-08-731-722-5	Sequence 5, Appli
	28	51.2	15.7	2230	4	US-08-378-313-24	Sequence 24, Appl
	29	50.8	15.5	1477	3	US-08-096-181A-7	Sequence 7, Appli
	30	50.8	15.5	1477	5	PCT-US94-08326-7	Sequence 7, Appli
	31	50.6	15.5	860	4	US-08-998-416-287	Sequence 287, App
	32	50.6	15.5	2110	4	US-09-419-459-1	Sequence 1, Appli
c	33	50.4	15.4	1431	4	US-09-316-083-2	Sequence 2, Appli
c	34	50.2	15.4	1850	3	US-08-617-860B-32	Sequence 32, Appl
c	35	50.2	15.4	4098	2	US-08-605-106-4	Sequence 4, Appli
	36	50	15.3	660	1	US-07-991-867B-32	Sequence 32, Appl
	37	50	15.3	660	1	US-08-107-755A-32	Sequence 32, Appl
	38	50	15.3	660	2	US-08-544-332-32	Sequence 32, Appl
c	39	50	15.3	734	3	US-09-014-583-1	Sequence 1, Appli
	40	50	15.3	1511	1	US-07-991-867B-8	Sequence 8, Appli
	41	50	15.3	1511	1	US-08-107-755A-8	Sequence 8, Appli
	42	50	15.3	1511	2	US-08-544-332-8	Sequence 8, Appli
c	43	50	15.3	1667	1	US-08-485-284A-1	Sequence 1, Appli
	44	50	15.3	4810	3	US-08-852-629-11	Sequence 11, Appl
	45	50	15.3	4838	3	US-08-852-629-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-306-593-1
; Sequence 1, Application US/09306593
; Patent No. 6184018
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; APPLICANT: Ximenes, Eduardo A.
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from
; TITLE OF INVENTION: Orpinomyces PC-2
; FILE REFERENCE: 31-98us
; CURRENT APPLICATION NUMBER: US/09/306,593
; CURRENT FILING DATE: 1999-05-06
; EARLIER APPLICATION NUMBER: US 60/084,494
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2435
; TYPE: DNA
; ORGANISM: Orpinomyces sp. PC-2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (39)..(2009)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (87)..(2009)
US-09-306-593-1

Bad date

Query Match 19.5%; Score 63.8; DB 4; Length 2435;
Best Local Similarity 56.9%; Pred. No. 0.0044;
Matches 136; Conservative 0; Mismatches 102; Indels 1; Gaps 1;

```
Qy     43 aaataaagataataaaaagatatattaattatatagctgaaaatttataatta-tatgat 101
          ||||||| ||| ||| ||   | |   |||| | |       ||||| || | |   ||
Db   2052 aaataaatatattaataatggaattattttattcacttcttttgctataagtagtgaaat 2111

Qy    102 aagtatagttaataaataaaaagtgttctcgggggacacttttttgttttaaaaaggaaa 161
          || |  | || |||| || | |   || | |     | ||||||||  | |  ||    |
Db   2112 aaattaattttataattatataaatttatagaataaatctttttttgaatcattaaaatta 2171

Qy    162 atataaataaaatttagataaaagtgtaaaataattatttttatttttaaatttgttaaaa 221
          | |||||||| ||  | ||   | || | |||||| ||| |||||  | |||    | ||
Db   2172 aaataaataatatacaaattttaatgaataataatgattattattaaatattctaaagaa 2231

Qy    222 atttgatataattgaattgtaaaaaaaatttcaggggggaatataaatgaaaaaaatta 280
            || |||   || ||  |||| | ||    ||     ||    ||| ||||||| ||
Db   2232 gatttataatttttaagaataaatataaagcaagaaaacaaatataattaaaaaaaata 2290
```

```
RESULT   2
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Eric A.
;    APPLICANT:  Bradshaw, Marite
;    APPLICANT:  Rood, Julian
;    TITLE OF INVENTION:  Expression System for Clostridium
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,075
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  960296.95238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3770..4013
;      OTHER INFORMATION:  /note= "RP4 origin of DNA transfer (oriT) from
;      OTHER INFORMATION:  plasmid RP4"
US-09-056-075-1
```

Bad Date

```
  Query Match           19.0%;  Score 62.2;  DB 2;  Length 6243;
  Best Local Similarity   49.5%;  Pred. No. 0.0071;
  Matches  160;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
Db    1183 ATGAGGGTAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 1242

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
Db    1243 AATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAATA 1302

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
Db    1303 TAAAAATAAAAAAATATAAAAATAAAAAAATATAAAAATAAAAAAATATAAAAATAAAAA 1362

Qy     181 aaaagtgtaaaataattatttttatttaaatttgttaaaaatttgatataattgaattg 240
Db    1363 AATAAAAAAATATAAAAATAAAAAAATAAAAAAATATAAAAATATTTTTTATTTAAAGTT 1422

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
```

```
Db    1423 TGAAAAAAATTTTTTTATATTATATAATCTTTGAAGAAAAGAATATAAAAAATGAGCCTT 1482

Qy     301 tttatgttttcatgttttcttat 323
Db    1483 TATAAAAGCCCATTTTTTTTCAT 1505


RESULT   3
US-08-998-416-186/c
; Sequence 186, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  186:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1074RP
US-08-998-416-186
```

```
  Query Match           18.9%;  Score 61.8;  DB 4;  Length 615;
  Best Local Similarity   52.1%;  Pred. No. 0.01;
  Matches  138;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy      62 atatattaattatatagctgaaaatttataattatatgataagtatagttaataaataaa 121
Db     596 ATTAATTAATTAAATAATGATAAAATATTAATTAAATAATTAATTTAATTAAATAAATAA 537

Qy     122 aagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagata 181
Db     536 ATAAATAAAATGAATATAATTTATTAAATAATATAAAATTAATAATAAGAAATTAAAGT 477

Qy     182 aaagtgtaaaataattatttttatttaaatttgttaaaaatttgatataattgaattgt 241
```

```
Db    476 TAAAATTAATTTTAATAATAATTCTTATAAAAAGTTAAATAATATAAATCAACATAATAT 417

Qy    242 aaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaattt 301
Db    416 TTATAAAAATAGATATTATAATAAAAATAATAATTTACAATATTTAAATAAATTAAATCT 357

Qy    302 ttatgttttcatgttttcttattgt 326
Db    356 TTATAATAATAAAATATATTATTTT 332


RESULT   4
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match           18.8%;  Score 61.6;  DB 2;  Length 19124;
  Best Local Similarity   51.1%;  Pred. No. 0.0079;
  Matches  145;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy      8 atatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaagatatat 67
Db  15743 ATATCATAAAATAAAAAAAAATTAAAAAAATGTTAAAAAAAAAATATATACATAAAATAA 15802

Qy     68 taattatatagctgaaaatttataattatatgataagtatagttaataaataaaaagtgt 127
Db  15803 AAAAAATTTATTTAAATAAAAAAAAATAATAAATAAAAAAATTTAATTAAATAAAAAAA 15862

Qy    128 tctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagataaaagtg 187
Db  15863 ATAATAAATAAAAAAATTTAATTAAATAAAAAAAAATTAAAAAAATTTAATGAAATAAAA 15922

Qy    188 taaaataattatttttattttaaatttgttaaaaatttgatataattgaattgtaaaaaa 247
Db  15923 AAAAATAAAAAAATTTAATTAAATAAAAAAAATAAAATAAAATTAATTACATGCACATAT 15982

Qy    248 aatttcaggggggaatataaatgaaaaaaattatttcaaagttt 291
Db  15983 ACATACATATATATATATATACCCATAACTACATACAACATT 16026


RESULT   5
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match           18.1%;  Score 59.2;  DB 1;  Length 5852;
  Best Local Similarity   49.7%;  Pred. No. 0.021;
  Matches  151;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy     24 acagaagaatgtttaaatgaaataaagataataaaaagatatattaattatatagctgaa 83
```

```
Db   1878 AAAAAAAAAATTTTAAAATTTAAAAAAAAAAGAAAAAAAAAAAAAAAAGTAGAATTTATTA 1819

Qy     84 aatttataattatatgataagtatagttaataaataaaaagtgttctcgggggacacttt 143
          ||  |  || |||     ||   || | ||| ||  | |  |      |       | |||
Db   1818 AAAATTAAAATATTATTCAATCTTAATAAATTAAGTATATATCGATAGGCAATTTATTTT 1759

Qy    144 tttgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaataattatttttt 203
          | | | |    ||  |||| ||   ||||| |  | |     ||  ||  | |  |||
Db   1758 TATATCTATCTAAAAAAAAACTAGGAAAAATGAATGTCATCAAATAGTATTTTAACATTT 1699

Qy    204 attttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaat 263
           ||||   ||| |||||||    | | ||   |   ||||||| | ||| |    ||
Db   1698 TTTTTTTTTTTTTTAAAAAAAAAGTGTCATGACAAAAAAAAAAAAGTGTCATGACAAAAA 1639

Qy    264 ataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttat 323
          | |||  ||||||       |    |||     |  ||  | ||||    ||||   ||
Db   1638 AAAAAAAAAAAAAGAGGGGAAAGTAATTATAACTAGGTTAGTTTTTTATAATTTTTACAT 1579

Qy    324 tgtt 327
            ||
Db   1578 ATTT 1575
```

```
RESULT   6
US-08-973-273-4/c
; Sequence 4, Application US/08973273
; Patent No. 6140085
;  GENERAL INFORMATION:
;    APPLICANT:  Dean, Caroline
;    APPLICANT:  MacKnight, Richard C
;    APPLICANT:  Bancroft, Ian
;    APPLICANT:  Lister, Clare K
;    TITLE OF INVENTION:  Genetic Control of Flowering
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye P.C.
;      STREET:  1100 No. 6140085th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  United States of America
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/973,273
;      FILING DATE:  01-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/01332
;      FILING DATE:  03-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9511196.9
;      FILING DATE:  02-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ms Mary J Wilson
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  620-29
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9048 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Brassica
```

```
US-08-973-273-4


  Query Match           17.8%;  Score 58.2;  DB 3;  Length 9048;
  Best Local Similarity   53.5%;  Pred. No. 0.029;
  Matches  144;  Conservative    0;  Mismatches  123;  Indels    2;  Gaps    1;

Qy     35 tttaaatgaaataaagataataaaaagatatattaattatatagctgaaaatttataatt 94
          | | |||||| | |    | ||||||   || ||| |||   ||   ||| | |||  |
Db    452 TCTTAATGAATTTATATAATTAAAAACGAATTTTATTTAATCAGTGAAAATTATATGTTA 393

Qy     95 atatgataagtatagttaataaataaaaagtgttctcgggggacacttttttgttttaaa 154
             | |||   ||| || ||   |||| | | |  |       |        | |   ||
Db    392 TATTAATATAAATAATTGATTCTTAAATAATATAATAAAATTATCAAAAAATTTAAAAAT 333

Qy    155 aaggaaaatataaataaaatttagataaaagtgtaaaataattatttttattttaaattt 214
          |||  || | || ||| | |||       |  || |||  |  |||||||||| ||||  |
Db    332 AAGTTAATTCTATATATATTTTGTTGCTATCTGAAAACAA--TATTTTTATTATAAAAGT 275

Qy    215 gttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaatataaatgaaaa 274
          ||||  || ||||||    ||| |  |||  |||||   |       |   | ||||
Db    274 TAAAAAACATTTATATAAAGATATTTTCTAAACTATTTCTAAGATATGAGTGTTTTAAAA 215

Qy    275 aaattatttcaaagtttactgtaatttttt 303
          || |||   |||   | | | ||  ||||
Db    214 AATTTAACACAATAATAAGTATATATTTT 186
```

```
RESULT   7
US-08-998-416-1137/c
; Sequence 1137, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1137:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1692RP
US-08-998-416-1137


  Query Match             17.7%;  Score 57.8;  DB 4;  Length 636;
  Best Local Similarity   49.5%;  Pred. No. 0.044;
  Matches  149;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy      26 agaagaatgtttaaatgaaataaagataataaaaagatatattaattatatagctgaaaa 85
           | ||| || ||||| |  | |    ||||||      ||  | | | |   |    ||||
Db     632 ATAAGTATTTTTAACTACATTCTTTATAATATTTTATTAATTAATTAAATAATGATAAAA 573

Qy      86 tttataattatatgataagtatagttaataaataaaaagtgttctcgggggacacttttt 145
           | |  | | |     ||| |  | | ||||||||||| |        |      | || ||
Db     572 TATTAATTAAATAATTAATTTAATTAAATAAATAAATAAATAAAATGAATATAATTTATT 513

Qy     146 tgttttaaaaaggaaaatataaataaaatttagataaaagtgtaaaataattatttttat 205
             |  | ||  |  |   | |  ||||| |  | ||| | |   |||| ||  || |
Db     512 AAATAATATAAAAATTAATAATAAGAAATTAAAGTTAAAATTAATTTTAATAATAATTCT 453

Qy     206 tttaaatttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaatat 265
           | ||||    |||||  | |  | ||  |    | | |  | |||| |  |       ||||
Db     452 TATAAAAAGATTAAATAATATAAATCAACATAATATTTATAAAAATAGATATTATAATAA 393

Qy     266 aaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattg 325
           ||||  | ||     | |    |  | | |  ||| |||||  |   |   | | |||||
Db     392 AAATAATAATTTACAATATTTAAATAAATTAAATCTTTATAATAATAAAATATATTATTT 333

Qy     326 t 326
           |
Db     332 T 332


RESULT   8
US-08-883-795A-36/c
; Sequence 36, Application US/08883795A
; Patent No. 5985607
;  GENERAL INFORMATION:
;    APPLICANT:  Delcuve, Genevieve
;    APPLICANT:  Awang, Gregor
;    TITLE OF INVENTION:  Recombinant DNA Molecules and Expression
;    TITLE OF INVENTION:  Vectors for Tissue Plasminogen Activator
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BERESKIN & PARR
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/883,795A
;      FILING DATE:  27-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gravelle, Micheline
;      REGISTRATION NUMBER:  40,261
;      REFERENCE/DOCKET NUMBER:  7841-062
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 364-7311
;      TELEFAX:  (416) 361-1398
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  665 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Rh 32
US-08-883-795A-36


  Query Match             17.4%;  Score 56.8;  DB 2;  Length 665;
  Best Local Similarity   50.6%;  Pred. No. 0.062;
  Matches  164;  Conservative    0;  Mismatches  157;  Indels    3;  Gaps    1;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
           | |  | |  | | ||  ||    |    | ||| | ||| |  ||  |   |||| | ||
Db     353 AATATGTAATTATAAACATTTTAATTATAAAATATGTAATTATAAACATTTTAATTATAA 294

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
             |||| | | |   |    |  || | || |||  |    ||   ||| || |   ||||
Db     293 AATATTTAATTATAAACATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAA 234

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           ||  | | |          ||  || |   | |    |  ||| | ||   | ||| |
Db     233 AATATTTAATTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAA 174

Qy     181 aaaagtgtaaaataattatttttatttttaaatttgttaaaaatttgatataattgaattg 240
           ||.| |  |  |||| |||||| ||| |||| |  ||   ||||  | ||| || ||||
Db     173 AATATTTAATTATAAATATTTTAATTATAAAATATTT---AATTATAAATATTTTAATTA 117

Qy     241 taaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgtaatt 300
           ||||| |  |          | | ||| |  |||| |||   | | |  |   |  | |
Db     116 TAAAATATTTAATTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTA 57

Qy     301 tttatgttttcatgttttcttatt 324
           |  || |||| ||  |    ||||
Db      56 TAAATATTTTAATTATAAAATATT 33


RESULT   9
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
```

```
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  288:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  837 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1241RP
US-08-998-416-288


  Query Match             17.2%;  Score 56.4;  DB 4;  Length 837;
  Best Local Similarity   52.1%;  Pred. No. 0.07;
  Matches  149;  Conservative    0;  Mismatches  136;  Indels    1;  Gaps    1;

Qy      38 aaatgaaataaagataataaaaagatatattaattatatagctgaaaatttataattata 97
           | || ||||||    ||| || | |    |||   | |||| |   |   || || ||
Db     445 AGATTAAATAATATAAATCAACATAATATTTATAAAAATAGATATTATAATAAAAATAAT 386

Qy      98 tgataagtatagttaataaataaaaagtgttctcgggggacacttttttgttttaaaaag 157
             | |  ||| ||||  ||   |||   || |     | | | | || |||||| ||
Db     385 AATTTACAATATTTAAATAAATTAAATCTTTATAATAATAAAATATATTATTTTAATAAA 326

Qy     158 gaaaatataaataaaatttagataaaagtgtaaaataattatttttattttaaatttgtt 217
             ||   |||||||| || |  || |  |  |||  || ||| |  | | |||||    |
Db     325 CAATTAATAAATAATATATTAATTATTGA-TAATCTATTTAATAATTTATTAAAGAAAAT 267

Qy     218 aaaaatttgatataattgaattgtaaaaaaaatttcaggggggaatataaatgaaaaaaa 277
           || ||| |   |||||    | |||  ||  | . |   |   ||| |    | | |
Db     266 AATAATATCTAATAATATTTTAATAACTAATTTAAAATTTGAACATAGACTAAATAGTAT 207

Qy     278 ttatttcaaagtttactgtaatttttatgttttcatgttttcttat 323
           | || | |||  ||| | | ||  |||| |    ||   |  | ||
Db     206 TCATATTAAATATTATTTTTATAATTATATAAATATTAATAATGAT 161


RESULT  10
US-08-617-860B-33
; Sequence 33, Application US/08617860B
; Patent No. 6133506
;  GENERAL INFORMATION:
;    APPLICANT:  Tvpfer, R., Bautor, J., Bothmann, H., Filsak, E.,
;    APPLICANT:  Hvricke-Grandpierre, C., Klein, B., Martini, N.,
;    APPLICANT:  M ller, A., Schulte, W., Voetz, M., Walek, J.,
;    APPLICANT:  Schell, J.
;    TITLE OF INVENTION:  Promoters
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steinberg, Raskin & Davidson, P.C.
;      STREET:  1140 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,860B
;      FILING DATE:  01-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02950
;      FILING DATE:  05-SEP-1994
;      APPLICATION NUMBER:  DE P4329951.2
;      FILING DATE:  04-SEP-1993
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2750 Base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double  stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Cuphea lanceolata
;    IMMEDIATE SOURCE:
;      LIBRARY:  genomic Lambda FIX II
;      CLONE:  C1TEg4
;    FEATURE:
;      NAME/KEY:  Startcodon
;      LOCATION:  2637..2639
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2637..2750
US-08-617-860B-33


  Query Match             17.1%;  Score 55.8;  DB 3;  Length 2750;
  Best Local Similarity   53.2%;  Pred. No. 0.077;
  Matches  143;  Conservative    0;  Mismatches  122;  Indels    4;  Gaps    1;

Qy      34 gtttaaatgaaataaagataataaaaagatatattaattatatagctgaaaatttataat 93
           | |||||| ||||||| | | | |   || |||| ||| ||||||   | |||||||| |
Db    1505 GATTAAATAAAATAAAAAGACTTAGTCGAAATATAAATAATATAGTACACAATTTATACT 1564

Qy      94 tatatgataagtatagttaataaataaaaagtgttctcgggggacacttttttgttttaa 153
             | |  |||||| |   ||   | ||||    || |   ||       ||| | |
Db    1565 ATTTTATTAAGTAAAAAAAAACTAAAAAAGTATTTGTAATTGGTTTTCAATTTTTCTCTT 1624

Qy     154 aaaggaaaatataaataaaatttagataaaagtgtaaaataatt----atttttatttta 209
             ||   ||||    | ||||| || |    | |  ||| ||| |    |  |  ||||
Db    1625 CAATTGAAATTATTACAAAATATATAATTTAATAAAAATTAAATTTAAAGATAAATTTGT 1684

Qy     210 aatttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaatataaat 269
           ||||| || || | ||    ||||| | || |     ||  || |       | | ||
Db    1685 AATTTTTTTAATACTTCGACTAATTTATTTTTCGTTTAATATTAAAAATTTTAGTTGAAA 1744

Qy     270 gaaaaaaattatttcaaagtttactgtaa 298
           | | || | || |    | |||||
Db    1745 AGGGCATAATAATCCATAAAACATTGTAA 1773


RESULT  11
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
;  GENERAL INFORMATION:
;    APPLICANT:  Delcuve, Genevieve
;    APPLICANT:  Awang, Gregor
;    TITLE OF INVENTION:  Recombinant DNA Molecules and Expression
;    TITLE OF INVENTION:  Vectors for Tissue Plasminogen Activator
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BERESKIN & PARR
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/883,795A
;      FILING DATE:  27-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gravelle, Micheline
;      REGISTRATION NUMBER:  40,261
;      REFERENCE/DOCKET NUMBER:  7841-062
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 364-7311
;      TELEFAX:  (416) 361-1398
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  665 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Rh 32
US-08-883-795A-36


  Query Match             17.0%;  Score 55.6;  DB 2;  Length 665;
  Best Local Similarity   50.3%;  Pred. No. 0.096;
  Matches  165;  Conservative    0;  Mismatches  159;  Indels    4;  Gaps    1;

Qy       1 atttgggatatcttaaatttagcacagaagaatgtttaaatgaaataaagataataaaaa 60
          || |   |||  | |||| |     |  | ||| |||| |   ||  |   ||  |  ||
Db      34 ATATTTTATAATTAAAATATTTATAATTAAAATATTTATAATTAAATATTTTATAATTAA 93

Qy      61 gatatattaattatatagctgaaaatttataattatatgataagtatagttaataaataa 120
           |||| |  | |  |    |  | | ||| |||  |   |    ||| || |||| |||
Db      94 AATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAA 153

Qy     121 aaagtgttctcgggggacacttttttgttttaaaaaggaaaatataaataaaatttagat 180
           ||  | |             || |   |   | |     || || | ||   ||    |
Db     154 AATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAA 213

Qy     181 aaaagtgtaaaataattatttttattttaaatttgttaa----aaatttgatataattga 236
           || | |   || ||| ||||||    |||||  |  |||     |||| |  |||||||  |
Db     214 AATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAA 273

Qy     237 attgtaaaaaaaatttcaggggggaatataaatgaaaaaaattatttcaaagtttactgt 296
           | |||  | ||      |      |||  |||||   | |||||  |       | | |
Db     274 AATGTTTATAATTAAATATTTTATAATTAAAATGTTTATAATTACATATTTTATAATTAA 333

Qy     297 aatttttatgttttcatgttttcttatt 324
           ||| |||||  || ||| |||| | |||
Db     334 AATGTTTATAATTACATATTTTATAATT 361


RESULT  12
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29.009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic
US-08-446-855A-1


  Query Match             16.8%;  Score 54.8;  DB 2;  Length 8920;
  Best Local Similarity   53.3%;  Pred. No. 0.099;
  Matches  163;  Conservative    0;  Mismatches  137;  Indels    6;  Gaps    2;

Qy      28 aagaatgtttaaatgaaataaagataataaaaagatatattaattatatagctgaaaatt 87
           || |   |||| |     ||| | | ||| |    ||| ||  |||||      ||| ||
Db    8783 AAAAGAATTTATACATTTTAATGTTTATATATTTTTATTTTTTTTATAATTAATAAATTT 8724

Qy      88 tataattatatgataagtatagttaataaataaaaagtgttctcgggggacacttttttg 147
           |||   ||| | |    | || |||  | ||||    ||        |||  |   ||
Db    8723 ATTAAAATTATAAAATTATTTGTAAATTGAAAAAACCATTTTGGTTTATACATATGAGTG 8664

Qy     148 ttttaaaaa----ggaaaatataaataaaatttagataaaagtgtaaaataattatttttt 203
           | || ||     | | |||||| |||||||| | |||| |   |  | || | ||
Db    8663 AATAAATAAAATTTGTATATATAACAAAAATTTATAAAAAAATAAATCAATATAAATTAA 8604

Qy     204 attttaaatttgttaaaa--atttgatataattgaattgtaaaaaaaatttcaggggga 261
           | | ||||  |||||  ||||  ||  ||| ||  ||||| | ||||  |      |
Db    8603 TGATCATATTTAATAAAAACATTTATTAATATTTAAACGTAAATATAATTAAATAAATAA 8544

Qy     262 atataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttctt 321
           ||||| ||  | | |  ||| |  |  ||   | | || | |   | || || | ||  |
Db    8543 ATATATATATATATATATATATATATTTATATTATATATTACTATTTATATATTTAT 8484

Qy     322 attgtt 327
           | | ||
Db    8483 ACTTTT 8478


RESULT  13
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
;  APPLICANT: Stewart et al.
;  TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
;  TITLE OF INVENTION: Synthetase II
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/150,741
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PL6380
;  EARLIER FILING DATE: 1992-12-16
;  EARLIER APPLICATION NUMBER: AU93/00617
;  EARLIER FILING DATE: 1993-12-02
;  EARLIER APPLICATION NUMBER: 08/446,855
;  EARLIER FILING DATE: 1995-07-06
;  NUMBER OF SEQ ID NOS: 15
```

```
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 8920
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-150-741-1


  Query Match             16.8%;  Score 54.8;  DB 4;  Length 8920;
  Best Local Similarity   53.3%;  Pred. No. 0.099;
  Matches  163;  Conservative    0;  Mismatches  137;  Indels    6;  Gaps    2;

Qy      28 aagaatgtttaaatgaaataaagataataaaaagatatattaattatatagctgaaaatt 87
           || |   |||| |     ||| | | ||| |    ||| ||  |||||      ||| ||
Db    8783 AAAAGAATTTATACATTTTAATGTTTATATATTTTTATTTTTTTTATAATTAATAAATTT 8724

Qy      88 tataattatatgataagtatagttaataaataaaaagtgttctcgggggacacttttttg 147
           |||   ||| | |    | || |||  | ||||    ||        |||  |   ||
Db    8723 ATTAAAATTATAAAATTATTTGTAAATTGAAAAAACCATTTTGGTTTATACATATGAGTG 8664

Qy     148 ttttaaaaa----ggaaaatataaataaaatttagataaaagtgtaaaataattatttttt 203
           | || ||     | | ||||||  |||||||| | |||| |  |  |  || | ||
Db    8663 AATAAATAAAATTTGTATATATAACAAAAATTTATAAAAAAATAAATCAATATAAATTAA 8604

Qy     204 attttaaatttgttaaaa--atttgatataattgaattgtaaaaaaaatttcaggggggа 261
           | | ||||  |||||  ||||  ||  ||| ||  ||||| | ||||  |      |
Db    8603 TGATCATATTTAATAAAACATTTATTAATATTTAAACGTAAATATAATTAAATAAATAA 8544

Qy     262 atataaatgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttctt 321
           ||||| ||  | | |  ||| |  |  ||   | | || | |   | || || | ||  |
Db    8543 ATATATATATATATATATATATATATTATATATTATATATTACTATTTATATATTTAT 8484

Qy     322 attgtt 327
           | | ||
Db    8483 ACTTTT 8478


RESULT  14
US-08-991-677-11
; Sequence 11, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
;  APPLICANT: Chiang, Vincent L
;  APPLICANT: Carraway, Daniel T
;  APPLICANT: Smeltzer, Richard H
;  TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
;  FILE REFERENCE: 50617
;  CURRENT APPLICATION NUMBER: US/08/991,677A
;  CURRENT FILING DATE: 1997-12-16
;  EARLIER APPLICATION NUMBER: US 60/033,381
;  EARLIER FILING DATE: 1996-12-16
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 2251
;   TYPE: DNA
;   ORGANISM: Pinus taeda
US-08-991-677-11


  Query Match             16.7%;  Score 54.6;  DB 4;  Length 2251;
  Best Local Similarity   53.3%;  Pred. No. 0.12;
  Matches  137;  Conservative    0;  Mismatches  119;  Indels    1;  Gaps    1;

Qy      32 atgtttaaatgaaataaagataataaaaagatatattaattatatagctgaaaatttata 91
           |||||     |   | || |  ||| | | ||||||  ||| |   | |||   |  ||
Db    536 atgttatggagtttttaaaaatatatatatatatatatttttgggttgagtttactta 595

Qy      92 attatatgataagtatagtta-ataaataaaaagtgttctcgggggacacttttttgttt 150
           |   |  || |||  | || | |   |||||  | ||| |     |     ||| | | ||
Db    596 aaatttggaaaaggttggtaagaactataaattgagttgtgaatgagtgttttatggatt 655

Qy     151 taaaaaggaaaatataaataaaatttagatataaaagtgtaaaataattatttttattttaa 210
           |   |||      |      | || ||  ||||| | ||   | | ||     | ||  |
Db    656 ttttaagatgttaaatttatatatgtaattaaaatttttattttgaataacaaaaattata 715

Qy     211 atttgttaaaaatttgatataattgaattgtaaaaaaaatttcaggggggaatataaatg 270
           ||| | |||||| ||| | |  |  | ||  |  ||||||||||      || |||| |
Db    716 attggataaaaaattgttttgttaaatttagagtaaaaatttcaaaatctaaaataatta 775

Qy     271 aaaaaaattatttcaaa 287
           || |  |||||||  ||
Db    776 aacactattatttttaa 792


RESULT  15
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match             16.6%;  Score 54.4;  DB 1;  Length 5852;
  Best Local Similarity   51.7%;  Pred. No. 0.12;
  Matches  124;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy      40 atgaaataaagataataaaaagatatattaattatatagctgaaaatttataattatatg 99
           ||||||  | | |||| ||| | |  |  | |   |||     |||||| || ||  | |
Db    176 ATGAAAAAGACATAAAAAATAAAGGTCATCAAAGTATTAAAAAAAATTAATTATCTTTTT 235
```

```
Qy     100 ataagtatagttaataaataaaaagtgttctcgggggacacttttttgttttaaaaagga 159
           |  .  |  |   || |||||||||   |        | | || ||||||||||| ||
Db     236 AACTTTGAAAAAAAAAAATAAAAAAAAATAAAAAAAAAAAAATTCTTTGTTTTAATAACTT 295

Qy     160 aaatataaataaaatttagataaaagtgtaaaataattatttttatttaaatttgttaa 219
             | |   || |||  |||   | | |  || || |  ||||||  |    |  |    |
Db     296 TTAAAATTATTAAAAATAGTATAGATTTAAAGATCACAATTTTTTATAATTAACTACATA 355

Qy     220 aaatttgatataattgaattgtaaaaaaaatttcaggggggaatataaatgaaaaaaatt 279
           ||||||   | || |||     |  || |  |  |       || ||| |   ||| |||
Db     356 AAATTTATAAAAAATGAGGGTCATGAAGATATATAAATAATTATTTAATTATTAAATATT 415
```

Search completed: January 24, 2002, 02:18:46
Job time: 3919 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:27:08 ; Search time 222.28 Seconds
(without alignments)
1261.226 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 989696

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16	4.9	29	17	AAT42655	Primer for amplify
2	16	4.9	29	21	AAA51200	N-terminal primer
3	16	4.9	32	20	AAZ27689	PCR primer for Ver
4	16	4.9	36	18	AAT97603	Shigella dysenteri
5	16	4.9	39	20	AAV90775	Primer Y104F. Syn
6	16	4.9	47	21	AAZ68379	Human map-related
7	15	4.6	20	21	AAA66183	Dog genomic marker
8	15	4.6	24	22	AAH46804	Human high motilit
9	15	4.6	27	17	AAT39439	Hel-N2 selected se
10	15	4.6	27	19	AAV37457	Human Hel-N2 selec
11	15	4.6	30	14	AAQ43975	Triple helix formi
12	15	4.6	48	16	AAT25576	Human gene signatu
13	14	4.3	18	13	AAQ20161	Cross-linking olig
14	14	4.3	18	13	AAQ20160	Cross-linking olig
15	14	4.3	18	13	AAQ30311	Oligomer HSV724 fo
16	14	4.3	18	13	AAQ30310	Oligomer HSV723 fo
c 17	14	4.3	18	20	AAZ22162	Human c-IAP-1 mRNA
18	14	4.3	22	22	AAS01590	Human IQGAP2 CpG i
c 19	14	4.3	22	22	AAS01643	Human IQGAP2 5' sta
c 20	14	4.3	26	19	AAV07952	Helicobacter pylor
c 21	14	4.3	26	19	AAV07922	Helicobacter pylor
c 22	14	4.3	27	19	AAV07937	Helicobacter pylor
c 23	14	4.3	29	17	AAT42655	Primer for amplify
c 24	14	4.3	29	21	AAA51200	N-terminal primer
c 25	14	4.3	31	16	AAT25703	Human gene signatu
26	14	4.3	31	19	AAV67854	Nucleotide fragmen
c 27	14	4.3	32	20	AAZ27689	PCR primer for Ver
c 28	14	4.3	36	18	AAT97603	Shigella dysenteri
29	14	4.3	36	22	AAC90606	Tomato spotted wil
30	14	4.3	37	15	AAQ62952	Glycophorin antibo
c 31	14	4.3	45	22	AAF55449	Oligonucleotide us
32	14	4.3	45	22	AAF55450	Oligonucleotide us
c 33	14	4.3	45	22	AAC88874	Oligonucleotide TA
34	14	4.3	45	22	AAC88875	Oligonucleotide TA
35	14	4.3	47	21	AAZ66366	Human map-related
c 36	14	4.3	47	21	AAZ67473	Human map-related
c 37	14	4.3	47	21	AAZ67533	Human map-related
c 38	14	4.3	47	21	AAZ67549	Human map-related
39	14	4.3	47	21	AAZ67813	Human map-related
40	14	4.3	50	21	AAA98312	Human MSH6 fragmen
41	13	4.0	15	22	AAF48097	IGFBP3 oligonucleo
42	13	4.0	15	22	AAF48098	IGFBP3 oligonucleo
43	13	4.0	15	22	AAF48099	IGFBP3 oligonucleo
c 44	13	4.0	16	21	AAA57758	Nucleotide sequenc
c 45	13	4.0	17	16	AAQ92084	Renilla reniformis

ALIGNMENTS

RESULT 1
AAT42655
ID AAT42655 standard; DNA; 29 BP.
XX
AC AAT42655;
XX
DT 25-FEB-1997 (first entry)
XX
DE Primer for amplifying verotoxin (VT-1) subunit A coding sequence.
XX
KW Verotoxin; Escherichia coli; enteric infection; diarrhoea; vaccine;
KW haemolytic uraemic syndrome; detection; ss.
XX
OS Synthetic.
XX
PN WO9630043-A1.
XX
PD 03-OCT-1996.
XX
PF 25-MAR-1996; 96WO-US04093.
XX
PR 24-MAR-1995; 95US-0410058.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Carroll SB, Padhye NV, Stafford DC;
XX
DR WPI; 1996-505779/50.
XX
PT Compsn. contg. neutralising antitoxin against E.coli vero-toxin -
PT used to treat intoxicated individuals, and as a prophylactic against
PT diarrhoeal disease or extra-intestinal complications of E.coli
PT infection
XX

PS Example 6; Page 58; 101pp; English.
XX
CC Compositions containing neutralising antitoxin against one or more E.
CC coli verotoxin (VT) can be used to treat intoxicated adults and
CC children with enteric bacterial infections. They may also be used as
CC prophylactics e.g. as a vaccine, against diarrhoeal disease or the
CC development of extra-intestinal complications of E.coli infection,
CC especially haemolytic uraemic syndrome. The antitoxin can also be
CC used to detect E. coli VT in a sample. The VT is recombinant,
CC preferably a fusion protein containing a non-VT protein sequence and
CC part of the E.coli VT1 or VT2 sequence. Two primers (AAT42655,
CC AAT42656) were used to amplify the verotoxin VT-1 A subunit coding
CC sequence and add a histidine tag coding sequence to the subunit
CC sequence. Two primers (AAT42655, AAT42658) were used to amplify the
CC verotoxin VT-1 A and B subunits and add a histidine tag coding
CC sequence to the subunit sequences.
XX
SQ Sequence 29 BP; 11 A; 2 C; 5 G; 11 T; 0 other;

Query Match 4.9%; Score 16; DB 17; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 aaaataattattttta 204
||||||||||||||||
Db 9 aaaataattattttta 24

RESULT 2
AAA51200
ID AAA51200 standard; DNA; 29 BP.
XX
AC AAA51200;
XX
DT 26-SEP-2000 (first entry)
XX
DE N-terminal primer for E. coli verotoxin 1 subunit A gene.
XX
KW VT-1; verotoxin; antitoxin therapy; fusion protein; affinity tag; food;
KW recombinant production; screening; dairy; anti-bacterial; vaccine;
KW primer; polyhistidine; ss.
XX
OS Escherichia coli.
OS Synthetic.
XX
PN US6080400-A.
XX
PD 27-JUN-2000.
XX
PF 13-MAR-1997; 97US-0816977.
XX
PR 24-MAR-1995; 95US-0410058.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Byrne LM;
XX
DR WPI; 2000-451195/39.
XX
PT Bacterial cell for recombinantly expressing bacterial toxins in large
PT quantities useful for immunization and treatment of bacterial
PT infections, comprises expression vector encoding bacterial toxin
XX
PS Example 6; Column 83; 83pp; English.
XX
CC E. coli verotoxin (VT) type 1 and 2 subunits A and B were cloned into
CC pET-23b, designed to allow expression of the native proteins containing
CC C-terminal polyhistidine tags. The VT-1 and VT-2 genes were engineered
CC to convert the signal sequence methionine codon into a NdeI site to
CC allow cloning of the amplified genes into the vector without addition of
CC vector-encoded amino acids. The C-terminal primers comprises the
CC C-terminal 7 codons of each gene fused to the sequence CTCGAGCC, in order
CC to add the polyhistidine tag. The primers delete the native stop codons,
CC and when cloned into pET-23 add a C-terminal extension of Leu-Glu-(His)6.
CC VT B chains are small proteins (approximately 8 kDa), so use of a small
CC affinity tag was preferred (i.e. polyhistidine). A polyhistidine affinity
CC tag facilitates single step affinity purification of subunits from
CC periplasmic extracts. However, due to poor recovery of his-tagged VT-1 A
CC and VT-2 A chains, expression of maltose binding protein (MBP) fused
CC subunits was undertaken. Due to the toxicity of the VT-2 B subunit,
CC strict uninduced promoter control is necessary to permit cell viability.
CC Bacterial host cells expressing a recombinant expression vector encoding
CC a polyhistidine affinity tag and a portion of the VT-2 B chain are
CC claimed. The vector is chosen from pET24hisVT2BL+, pET24hisVT2BL- and
CC pET24VT2B, where "L+" indicates that the vector encodes the preprotein
CC form of the protein and "L-" indicates that the vector encodes the mature
CC form of the protein. The bacterial cell is capable of expressing large
CC quantities (40 mg/l) of VT-2B. The toxins are useful for immunizing
CC non-mammals and for detecting bacterial toxins in environmental samples
CC including soil, water, industrial samples, biological samples and samples
CC obtained from food and dairy processing instruments.
XX
SQ Sequence 29 BP; 11 A; 2 C; 5 G; 11 T; 0 other;

Query Match 4.9%; Score 16; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 aaaataattattttta 204
||||||||||||||||
Db 9 aaaataattattttta 24

RESULT 3
AAZ27689
ID AAZ27689 standard; DNA; 32 BP.
XX
AC AAZ27689;
XX
DT 22-DEC-1999 (first entry)
XX
DE PCR primer for Verotoxin gene.
XX
KW Verotoxin; VT1; VT2; detection; PCR primer; ss.
XX
OS Synthetic.
OS Escherichia coli.
XX
PN JP11243996-A.
XX
PD 14-SEP-1999.
XX
PF 27-FEB-1998; 98JP-0047677.
XX
PR 27-FEB-1998; 98JP-0047677.
XX
PA (TOYM) TOYOBO KK.
XX
DR WPI; 1999-603716/52.
XX
PT An oligonucleotide for amplification of verotoxin - useful in the
PT detection of inactivated verotoxin gene by transfer of a foreign DNA
PT fragment
XX
PS Claim 11; Page 9; 10pp; Japanese.
XX
CC This sequence represents a PCR primer of the invention. The primer is
CC used for amplification of the E. coli verotoxin (VT) gene. The
CC oligonucleotide is useful for detection of inactivated VT gene by
CC transfer of a foreign DNA fragment. Simple, rapid and specific
CC amplification of VT gene from environmental factors is achieved using the
CC oligonucleotide of the invention.
XX
SQ Sequence 32 BP; 12 A; 2 C; 4 G; 14 T; 0 other;

```
  Query Match             4.9%;  Score 16;  DB 20;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattattttta 204
          ||||||||||||||||
Db      4 aaaataattattttta 19
```

```
RESULT   4
AAT97603
ID   AAT97603 standard; DNA; 36 BP.
XX
AC   AAT97603;
XX
DT   30-APR-1998  (first entry)
XX
DE   Shigella dysenteriae delta-stxA allele PCR primer 13.
XX
KW   Delta-virG allele; delta-guaB-A allele; PCR; amplification; primer;
KW   delta-stxA allele; shigellosis; vaccine; ss.
XX
OS   Synthetic.
OS   Shigella dysenteriae.
XX
PN   WO9737685-A1.
XX
PD   16-OCT-1997.
XX
PF   09-APR-1997;   97WO-US05954.
XX
PR   09-APR-1996;   96US-0629600.
XX
PA   (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI   Levine MM,  Noriega FR;
XX
DR   WPI; 1997-512417/47.
XX
PT   Shigella mutants with mutation in guaB-A - used in vaccines against
PT   Shigellosis
XX
PS   Example 6; Page 57; 94pp; English.
XX
CC   This is a PCR primer used in the amplification of the Shigella
CC   dysenteriae 1 delta-stxA allele.  The delta-stxA allele was integrated
CC   into delta-guaB-A of delta-guaB-A, delta-virG S. dysenteriae 1, which
CC   inactivated the shiga toxin of this strain.  The mutant can be used in
CC   the preparation of vaccines such as, a live vector vaccine comprising
CC   a Shigella mutant, (which encodes and expresses a foreign antigen, and
CC   a pharmaceutically acceptable carrier) or a DNA mediated vaccine
CC   comprising the Shigella mutant (which also contains a plasmid which
CC   encodes and expresses a foreign antigen in a eukaryotic cell).  The
CC   vaccines can be used against Shigellosis.
XX
SQ   Sequence 36 BP; 11 A; 3 C; 10 G; 12 T; 0 other;
```

```
  Query Match             4.9%;  Score 16;  DB 18;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattattttta 204
          ||||||||||||||||
Db     14 aaaataattattttta 29
```

```
RESULT   5
AAV90775
ID   AAV90775 standard; DNA; 39 BP.
XX
AC   AAV90775;
XX
DT   18-FEB-1999  (first entry)
XX
DE   Primer Y104F.
XX
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; primer; ss.
XX
OS   Synthetic.
XX
PN   WO9849314-A2.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-US08487.
XX
PR   14-OCT-1997;   97US-0061958.
PR   25-APR-1997;   97US-0045107.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim Disclosure; Page 194; 402pp; English.
XX
CC   The specification, which describes Helicobacter pylori antigenic
CC   proteins that are characterised by immunoreactivity with
CC   H. pylori-positive antisera. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori
CC   antigens are used to detect H. pylori-specific antibodies, for
CC   diagnosing infection or to confirm eradication of infection, and
CC   in vaccines to protect against H. pylori infection and related
CC   diseases (gastritis, peptic ulcer, gastric adenocarcinoma/lymphoma).
CC   The present primer is used in the course of the invention.
XX
SQ   Sequence 39 BP; 15 A; 8 C; 7 G; 9 T; 0 other;
```

```
  Query Match             4.9%;  Score 16;  DB 20;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 atatcttaaatttagc 23
          ||||||||||||||||
Db     23 atatcttaaatttagc 38
```

```
RESULT   6
AAZ68379
ID   AAZ68379 standard; DNA; 47 BP.
XX
AC   AAZ68379;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:2726.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,A)
```

```
FT                  /*tag=  a
FT                  /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 3; Page 812; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 24 A; 5 C; 6 G; 12 T; 0 other;


  Query Match             4.9%;  Score 16;  DB 21;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 9.8e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    147 gttttaaaaaggaaaa 162
          ||||||||||||||||
Db     13 gttttaaaaaggaaaa 28


RESULT   7
AAA66183
ID   AAA66183 standard; DNA; 20 BP.
XX
AC   AAA66183;
XX
DT   09-OCT-2000  (first entry)
XX
DE   Dog genomic marker oligonucleotide sequence SEQ ID NO:45.
XX
KW   Dog; genome; genomic marker; radiation hybrid map; identification;
KW   chromosome location; gene marker; polymorphic microsatellite marker;
KW   phenotype; behaviour; pedigree; ss.
XX
OS   Canis familiaris.
XX
PN   WO200029615-A2.
XX
PD   25-MAY-2000.
XX
PF   15-NOV-1999;   99WO-IB01907.
XX
PR   13-NOV-1998;   98US-0108193.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Galibert F,  Andre C;
XX
DR   WPI; 2000-387821/33.
XX
PT   New radiation hybrid map of the dog, Canine familiaris, genome, useful
PT   for e.g. identifying genes implicated in phenotypic and behavioral
PT   traits or in genetic diseases and for studying dog pedigrees  -
XX
PS   Claim 1; Page 55; 87pp; English.
XX
CC   The present invention describes a radiation hybrid map of the dog
CC   (Canine familiaris) genome comprising the genome location of a marker
CC   selected from AAA66139 to AAA66942. The radiation hybrid map is useful
CC   for identifying and localising dog genes, since it covers approximately
CC   80 % of the dog genome and provides a dense map integrating different
CC   types (i.e. Type I and Type II) of markers. The map and the dog genome
CC   markers (or complementary sequences) are especially useful to identify
CC   genes responsible for phenotypic and behavioural traits in dogs, to
CC   identify morbid genes, to analyse diseases and identify implicated genes
CC   in such diseases and their alleles, and to study dog pedigrees. They
CC   may also be useful for isolating corresponding human gene sequences
CC   e.g. genes involved in genetic diseases.
XX
SQ   Sequence 20 BP; 14 A; 1 C; 2 G; 3 T; 0 other;


  Query Match             4.6%;  Score 15;  DB 21;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    261 aatataaatgaaaaa 275
          |||||||||||||||
Db      1 aatataaatgaaaaa 15


RESULT   8
AAH46804
ID   AAH46804 standard; DNA; 24 BP.
XX
AC   AAH46804;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human high motility group protein family 11 cDNA PCR primer #2.
XX
KW   Human; high motility group protein family 11; cancer; haemopathy;
KW   HIV infection; immune disease; inflammation; gene therapy;
KW   PCR primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200147967-A1.
XX
PD   05-JUL-2001.
XX
PF   18-DEC-2000; 2000WO-CN00595.
XX
PR   23-DEC-1999;   99CN-0125721.
XX
PA   (UYFU-) UNIV FUDAN.
PA   (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2001-418226/44.
XX
PT   High motility group protein family 11 and encoded polynucleotide,
PT   applicable in diagnosis and treatment of cancer, haemopathy, HIV
```

```
PT   infection, immunological diseases and various inflammation  -
XX
PS   Example 3; Page 17; 39pp; Chinese.
XX
CC   The present invention provides the protein and coding sequences of the
CC   human high motility group protein family 11. The sequences are useful in
CC   the treatment of cancer, haemopathy, HIV infection, immune diseases and
CC   inflammation. The present sequence is a PCR primer for the coding
CC   sequence of the invention.
XX
SQ   Sequence 24 BP; 8 A; 3 C; 0 G; 13 T; 0 other;


  Query Match             4.6%;  Score 15;  DB 22;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    193 taattatttttattt 207
          ||||||||||||||||
Db      2 taattatttttattt 16


RESULT   9
AAT39439
ID   AAT39439 standard; RNA; 27 BP.
XX
AC   AAT39439;
XX
DT   13-NOV-1996  (first entry)
XX
DE   Hel-N2 selected sequence, e-13.
XX
KW   Human; neuron-specific protein; Hel-N1; 3'-UTR instability sequence;
KW   paraneoplastic sensory neuropathy; oncoprotein; lymphokine; rat; elav;
KW   RNA recognition motif; RRM; Drosophila; cellular growth; localisation;
KW   instability; translatability; neurons; autoimmune protein; PE; PCD; PSN;
KW   central nervous system; cancer; paraneoplastic cerebellar degeneration;
KW   paraneoplastic encephalomyelitis; RNP-1 octamer sequence;
KW   human elav-like neuronal protein-1; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        15..23
FT                   /*tag=  a
FT                   /note= "Consensus sequence"
XX
PN   US5525495-A.
XX
PD   11-JUN-1996.
XX
PF   11-MAY-1992;   92US-0881075.
XX
PR   15-SEP-1993;   93US-0120827.
PR   11-MAY-1992;   92US-0881075.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Gao F,  Keene JD,  Levine T;
XX
DR   WPI; 1996-286398/29.
XX
PT   Prodn of cDNA library for related proteins - by screening total cell
PT   mRNA with RNA-binding protein Hel-N1 or Hel-N2
XX
PS   Disclosure; Fig 11; 66pp; English.
XX
CC   The sequences given in AAT39427-46 represent random oligonucleotides
CC   which were isolated due to their ability to bind to the human neuron-
CC   specific protein, Hel-N2.  These sequences contain short stretches
CC   of uridylate residues interspersed with other residues.  These U-
CC   rich regions share homology with the 3'-UTR instability sequences
CC   that are found in mRNA's.  Instability sequences are target elements
CC   which reside in the 3'-non-coding regions of mRNA's which encode
CC   oncoproteins and lymphokines.  Hel-N2 is a deleted form of Hel-N1
CC   in which the residues 239-251 have been deleted.  This protein is
CC   expressed in medulloblastoma tumour cells and is not found in whole
CC   human brain.  A small amount of Hel-N2 is also found in fetal brain,
CC   which may indicate a correlation with rapid growth.  Hel-N1 cDNA was
CC   isolated by probing for rat and human elav counterparts using degenerate
CC   primers designed to simulate the RNP-1 octamer sequence present in two
CC   of the three RNA recognition motifs (RRM's) of Drosophila elav.  Hel-N1
CC   was found to contain 3 RRM's, where the third one (see also AAW00244) is
CC   sufficient for mRNA 3'-UTR binding activity.  Full length Hel-N1, when
CC   transfected into a cell, causes cellular growth to cease, however, if
CC   just the third binding domain is transfected into cells, the cells
CC   undergo rapid growth.  Hel-N1 binds as a multimer along the mRNA,
CC   presumably enhancing its localisation, instability and/or regulating it
CC   translatability and/or deadenylating it.  This protein may be
CC   responsible for the growth cessation of neurons. Hel-N1 is an autoimmune
CC   protein in certain patients who show central nervous system
CC   manifestations of cancer called paraneoplastic cerebellar degeneration
CC   (PCD), paraneoplastic encephalomyelitis (PE) or paraneoplastic sensory
CC   neuropathy (PSN).
XX
SQ   Sequence 27 BP; 7 A; 2 C; 2 G; 16 U; 0 other;


  Query Match             4.6%;  Score 15;  DB 17;  Length 27;
  Best Local Similarity   33.3%;  Pred. No. 2.6e+03;
  Matches    5;  Conservative   10;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    197 tatttttattttaaa 211
          :|:::::|::::|||
Db     12 uauuuuuauuuuaaa 26


RESULT  10
AAV37457
ID   AAV37457 standard; RNA; 27 BP.
XX
AC   AAV37457;
XX
DT   07-SEP-1998  (first entry)
XX
DE   Human Hel-N2 selected RNA sequence e-13.
XX
KW   Growth regulatory protein; Hel-N2; oncogene; cytokine; lymphokine;
KW   chromosome mapping; human; functionally related protein; ss.
XX
OS   Homo sapiens.
XX
PN   US5773246-A.
XX
PD   30-JUN-1998.
XX
PF   07-JUN-1995;   95US-0478675.
XX
PR   15-SEP-1993;   93US-0120827.
PR   11-MAY-1992;   92US-0881075.
PR   07-JUN-1995;   95US-0478675.
XX
PA   (GAOF/) GAO F.
PA   (KEEN/) KEENE J D.
PA   (LEVI/) LEVINE T.
XX
PI   Gao F,  Keene JD,  Levine T;
XX
DR   WPI; 1998-387003/33.
XX
PT   Use of proteins which bind RNA - for obtaining a cDNA library
PT   containing members encoding structurally or functionally related
PT   proteins from total cell mRNA.
XX
PS   Disclosure; Fig 11; 67pp; English.
XX
```

```
CC   This represents a RNA sequence selected from a randomised RNA library
CC   by binding to human Hel-N2 under high stringency conditions. The
CC   invention provides a method for obtaining a cDNA library having members
CC   encoding a group of structurally or functionally related proteins from
CC   total cell mRNA. The method comprises binding RNA representing total cell
CC   mRNA to a protein that binds RNA and has specific binding to untranslated
CC   regions of a subset of the total cell mRNA, where the protein that binds
CC   RNA has been purified to remove any other protein that binds RNA. The
CC   resulting bound products are separated and a cDNA library is prepared.
CC   The protein that binds RNA may be Hel-N1, Hel-N2, CArG, DT-7, K1, K2,
CC   K3, HuD, HuC, elav, rbp9, eIF4B, sxl, tra-2, AUBF, AUF, ASF/SF2, U2AF,
CC   SC35, or hnRNP proteins. The cDNAs obtained can be used for chromosome
CC   mapping and genome sequencing of the structurally or functionally related
CC   genes encoding growth regulatory proteins, proto-oncogenes, cytokines,
CC   lymphokines, or anti-oncogene proteins. They can also be used in
CC   diagnostic methods for determining imbalances in such genes and the
CC   efficacy of various treatments to correct any imbalance.
XX
SQ   Sequence 27 BP; 7 A; 2 C; 2 G; 16 U; 0 other;


  Query Match             4.6%;  Score 15;  DB 19;  Length 27;
  Best Local Similarity   33.3%;  Pred. No. 2.6e+03;
  Matches    5;  Conservative   10;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    197 tatttttattttaaa 211
          :|:::::|::::|||
Db     12 uauuuuuauuuuaaa 26


RESULT  11
AAQ43975
ID   AAQ43975 standard; DNA; 30 BP.
XX
AC   AAQ43975;
XX
DT   28-OCT-1993  (first entry)
XX
DE   Triple helix forming oligonucleotide III.
XX
KW   Purine; pyrimidine; tracts; intramolecular triplex; therapeutic;
KW   diagnostic; control; gene expresssion; mRNA synthesis suppression;
KW   ss.
XX
OS   Synthetic.
XX
PN   WO9312230-A.
XX
PD   24-JUN-1993.
XX
PF   11-DEC-1992;   92WO-US10792.
XX
PR   13-DEC-1991;   91US-0808452.
PR   21-JAN-1992;   92US-0826934.
XX
PA   (STRI ) SRI INT.
XX
PI   Jayasena SD,  Johnston BH;
XX
DR   WPI; 1993-214172/26.
XX
PT   New oligo:nucleotide(s) forming triple helix with target nucleic
PT   acid - contain purine and pyrimidine tracts in specific
PT   orientations, useful therapeutically or diagnostically e.g. for
PT   inactivating HIV RNA, etc.
XX
PS   Disclosure; Page 48; 101pp; English.
XX
CC   The sequence is that of an oligonucleotide, III, which is able to form
CC   a triple helix with a duplex nucleic acid (dsNA) contg. a target
CC   sequence which comprises at least one pyrimidine tract, and at least
CC   one adjacent purine tract. It is useful for therapeutic or
CC   diagnostic control of gene expression, e.g. suppression of mRNA
CC   synthesis from a target gene. A specified application is targetting
CC   of RNA in the HIV-1 genome. When appropriately labelled it may also
CC   be used as a probe. Attachment of cleavage agents caused permanent
CC   inactivation of the target by site-specific cleavage.
XX
SQ   Sequence 30 BP; 12 A; 0 C; 5 G; 13 T; 0 other;


  Query Match             4.6%;  Score 15;  DB 14;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    212 tttgttaaaaatttg 226
          |||||||||||||||
Db      5 tttgttaaaaatttg 19


RESULT  12
AAT25576
ID   AAT25576 standard; cDNA to mRNA; 48 BP.
XX
AC   AAT25576;
XX
DT   15-NOV-1996  (first entry)
XX
DE   Human gene signature HUMGS07754.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 1862; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 48 BP; 26 A; 5 C; 4 G; 13 T; 0 other;
```

```
  Query Match             4.6%;  Score 15;  DB 16;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    159 aaaatataaataaaa 173
          ||||||||||||||||
Db     17 aaaatataaataaaa 31


RESULT  13
AAQ20161
ID   AAQ20161 standard; DNA; 18 BP.
XX
AC   AAQ20161;
XX
DT   01-APR-1992  (first entry)
XX
DE   Cross-linking oligomer 724 to target Herpes Simplex Virus I.
XX
KW   deoxyribonucleic acid; major groove; HSV;
KW   inverted polarity region; covalent cross-linking group; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base    1
FT                    /*tag=  a
FT                    /mod_base= OTHER
FT                    /note= "N4N4-ethanocytosine"
FT   modified_base    2
FT                    /*tag=  b
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    3
FT                    /*tag=  c
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    4
FT                    /*tag=  d
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    5
FT                    /*tag=  e
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    7
FT                    /*tag=  f
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    9
FT                    /*tag=  g
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    10
FT                    /*tag=  h
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    11
FT                    /*tag=  i
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   misc_feature     12..18
FT                    /*tag=  j
FT                    /label= inverted_polarity_region
FT                    /note= "see comments"
FT   modified_base    13
FT                    /*tag=  k
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    14
FT                    /*tag=  l
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    15
FT                    /*tag=  m
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    17
FT                    /*tag=  n
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
XX
PN   WO9118997-A.
XX
PD   12-DEC-1991.
XX
PF   24-MAY-1991;   91WO-1003680.
XX
PR   14-JAN-1991;   91US-0640654.
PR   25-MAY-1990;   90US-0529346.
XX
PA   (GILE-) GILEAD SCIE INC.
XX
PI   Matteucci MD,  Krawczyk S;
XX
DR   WPI; 1992-007480/01.
XX
PT   New sequence-specific non-photo-activated crosslinking agents -
PT   bind to the major groove of duplex DNA and are esp. useful for
PT   treating latent infections e.g. HIV
XX
PS   Example 4; Page 29; 42pp; English.
XX
CC   This oligomer contains an inverted polarity region formed from an
CC   o-xyloso dimer synthon. Residues 11 and 12 are linked via an
CC   o-xyloso group (i.e. nucleotides that have xylose sugar linked via
CC   the o-xylene ring). The sequence is designed to target the Herpes
CC   Simplex virus I beginning at nucleotide 10996 and to covalently
CC   cross-link to it. See also AAQ20151-Q20160.
XX
SQ   Sequence 18 BP; 12 A; 1 C; 0 G; 5 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 13;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    159 aaaatataaataaa 172
          ||||||||||||||
Db      2 aaaatataaataaa 15


RESULT  14
AAQ20160
ID   AAQ20160 standard; DNA; 18 BP.
XX
AC   AAQ20160;
XX
DT   01-APR-1992  (first entry)
XX
DE   Cross-linking oligomer 723 to target Herpes Simplex Virus I.
XX
KW   deoxyribonucleic acid; major groove; HSV;
KW   inverted polarity region; covalent cross-linking group; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base    1
FT                    /*tag=  a
FT                    /mod_base= OTHER
FT                    /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base    2
FT                    /*tag=  b
FT                    /mod_base= OTHER
```

```
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   3
FT                   /*tag= c
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   4
FT                   /*tag= d
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   5
FT                   /*tag= e
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   7
FT                   /*tag= f
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   9
FT                   /*tag= g
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   10
FT                   /*tag= h
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   11
FT                   /*tag= i
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   misc_feature    12..18
FT                   /*tag= j
FT                   /label= inverted_polarity_region
FT                   /note= "see comments"
FT   modified_base   13
FT                   /*tag= k
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   14
FT                   /*tag= l
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   15
FT                   /*tag= m
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   17
FT                   /*tag= n
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
XX
PN   WO9118997-A.
XX
PD   12-DEC-1991.
XX
PF   24-MAY-1991;    91WO-1003680.
XX
PR   14-JAN-1991;    91US-0640654.
PR   25-MAY-1990;    90US-0529346.
XX
PA   (GILE-) GILEAD SCIE INC.
XX
PI   Matteucci MD,  Krawczyk S;
XX
DR   WPI; 1992-007480/01.
XX
PT   New sequence-specific non-photo-activated crosslinking agents -
PT   bind to the major groove of duplex DNA and are esp. useful for
PT   treating latent infections e.g. HIV
XX
PS   Example 4; Page 29; 42pp; English.
XX
CC   This oligomer contains an inverted polarity region formed from an
CC   o-xyloso dimer synthon. Residues 11 and 12 are linked via an
CC   o-xyloso group (i.e. nucleotides that have xylose sugar linked via
CC   the o-xylene ring). The sequence is designed to target the Herpes
CC   Simplex virus I beginning at nucleotide 10996 and to covalently
CC   cross-link to it. See also AAQ20151-Q20161.
XX
SQ   Sequence 18 BP; 13 A; 0 C; 0 G; 5 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 13;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    159 aaaatataaataaa 172
          ||||||||||||||
Db      2 aaaatataaataaa 15


RESULT  15
AAQ30311
ID   AAQ30311 standard; DNA; 18 BP.
XX
AC   AAQ30311;
XX
DT   07-DEC-1992  (first entry)
XX
DE   Oligomer HSV724 for forming triplex with HSV target duplex.
XX
KW   Herpes simplex virus I; AIDS; modified; HIV; RSV; HPV; malignancy;
KW   hepetitis; inflammation; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag= a
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N4 N4 ethanocytosine"
FT   modified_base   2
FT                   /*tag= b
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   3
FT                   /*tag= c
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   4
FT                   /*tag= d
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   5
FT                   /*tag= e
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   7
FT                   /*tag= f
FT                   /mod_base= OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   9
FT                   /*tag= g
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   10
FT                   /*tag= h
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   11
FT                   /*tag= i
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   13
FT                   /*tag= j
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
```

```
FT   modified_base   14
FT                   /*tag= k
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   15
FT                   /*tag= l
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   17
FT                   /*tag=  m
FT                   /mod_base= OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   misc_feature    12..18
FT                   /*tag=  n
FT                   /label= inverted_polarity_region
FT                   /note= "see comments"
FT   misc_feature    11.12
FT                   /*tag=  o
FT                   /note= "o-xyloso dimer synthon linkage"
XX
PN   WO9209705-A.
XX
PD   11-JUN-1992.
XX
PF   25-NOV-1991;    91WO-US08811.
XX
PR   23-NOV-1990;    90US-0617907.
PR   18-JAN-1991;    91US-0643382.
PR   08-APR-1991;    91US-0683420.
PR   17-APR-1991;    91US-0686544.
PR   17-APR-1991;    91US-0686546.
PR   17-APR-1991;    91US-0686547.
PR   27-SEP-1991;    91US-0766733.
XX
PA   (GILE-) GILEAD SCI INC.
XX
PI   Froehler B,  Krawczyk S,  Matteucci MD,  Milligan J;
XX
DR   WPI; 1992-217083/26.
XX
PT   New oligomers contg. modified bases - which form a triplex with
PT   G-C doublet in a DNA duplex, for treating and diagnosing HIV,
PT   hepatitis, herpes, malignancy and inflammation
XX
PS   Claim 12; Page 67; 77pp; English.
XX
CC   The synthetic oligomer is capable of forming a triplex at
CC   physiological pH with a purine rich target sequence by coupling
CC   into the major groove of the duplex.   The specific target sequence
CC   of this oligomer is a herpes simplex virus I duplex beginning at
CC   nucleotide 10996 contg. a purine-rich region concentrated on
CC   one chain of the duplex.  The oligomer, and others like it are useful
CC   in diagnosis and therapy of diseases characterised by specific DNA
CC   duplex targets, e.g. respiratory syncytial virus, HIV, hepatitis,
CC   herpes, malignant tumours and inflammation. The triple helices form
CC   under mild conditions thus assays may be carried  out without
CC   subjecting the test specimen to harsh conditions.  The oligomer
CC   contains an inverted polarity region formed from an o-xyloso
CC   dimer synthon.  The linking gp. is o-xyloso (nucleotides have the 3'
CC   positions of xylose sugars linked via the o-xylene ring).  Two
CC   nucleotides are coupled through a xylene residue to form the dimer
CC   synthon.  This additional modifications may render the oligomer stable
CC   to nuclease activity.  The oligomer is able to inhibit gene expression,
CC   as verified by in vitro systems.
CC   See also AAQ25452-25501 and AAQ30226-448.
XX
SQ   Sequence 18 BP; 12 A; 1 C; 0 G; 5 T; 0 other;

  Query Match             4.3%;  Score 14;  DB 13;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     159 aaaatataaataaa 172
           ||||||||||||||
Db       2 aaaatataaataaa 15


RESULT  16
AAQ30310
ID   AAQ30310 standard; DNA; 18 BP.
XX
AC   AAQ30310;
XX
DT   07-DEC-1992  (first entry)
XX
DE   Oligomer HSV723 for forming triplex with HSV target duplex.
XX
KW   Herpes simplex virus I; AIDS; modified; HIV; RSV; HPV; malignancy;
KW   hepetitis; inflammation; ss.
XX
OS   Synthetic..
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag= a
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   2
FT                   /*tag= b
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   3
FT                   /*tag= c
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   4
FT                   /*tag= d
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   5
FT                   /*tag= e
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   7
FT                   /*tag= f
FT                   /mod_base= OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   9
FT                   /*tag= g
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   10
FT                   /*tag= h
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   11
FT                   /*tag= i
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   13
FT                   /*tag= j
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   14
FT                   /*tag= k
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   15
FT                   /*tag= l
FT                   /mod_base=  OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
FT   modified_base   17
FT                   /*tag=  m
FT                   /mod_base= OTHER
FT                   /note= "OTHER= N6 methyl-8-oxo 2' deoxyadenine"
```

```
FT   misc_feature    12..18
FT                   /*tag= n
FT                   /label= inverted_polarity_region
FT                   /note= "see comments"
FT   misc_feature    11.12
FT                   /*tag= o
FT                   /note= "o-xyloso dimer synthon linkage"
XX
PN   WO9209705-A.
XX
PD   11-JUN-1992.
XX
PF   25-NOV-1991;   91WO-US08811.
XX
PR   23-NOV-1990;   90US-0617907.
PR   18-JAN-1991;   91US-0643382.
PR   08-APR-1991;   91US-0683420.
PR   17-APR-1991;   91US-0686544.
PR   17-APR-1991;   91US-0686546.
PR   17-APR-1991;   91US-0686547.
PR   27-SEP-1991;   91US-0766733.
XX
PA   (GILE-) GILEAD SCI INC.
XX
PI   Froehler B,  Krawczyk S,  Matteucci MD,  Milligan J;
XX
DR   WPI; 1992-217083/26.
XX
PT   New oligomers contg. modified bases - which form a triplex with
PT   G-C doublet in a DNA duplex, for treating and diagnosing HIV,
PT   hepatitis, herpes, malignancy and inflammation
XX
PS   Claim 12; Page 67; 77pp; English.
XX
CC   The synthetic oligomer is capable of forming a triplex at
CC   physiological pH with a purine rich target sequence by coupling
CC   into the major groove of the duplex.   The specific target sequence
CC   of this oligomer is a herpes simplex virus I duplex beginning at
CC   nucleotide 10996 contg. a purine-rich region concentrated on
CC   one chain of the duplex.  The oligomer, and others like it are useful
CC   in diagnosis and therapy of diseases characterised by specific DNA
CC   duplex targets, e.g. respiratory syncytial virus, HIV, hepatitis,
CC   herpes, malignant tumours and inflammation. The triple helices form
CC   under mild conditions thus assays may be carried  out without
CC   subjecting the test specimen to harsh conditions.  The oligomer
CC   contains an inverted polarity region formed from an o-xyloso
CC   dimer synthon.  The linking gp. is o-xyloso (nucleotides have the 3'
CC   positions of xylose sugars linked via the o-xylene ring).  Two
CC   nucleotides are coupled through a xylene residue to form the dimer
CC   synthon.  This additional modifications may render the oligomer stable
CC   to nuclease activity.  The oligomer is able to inhibit gene expression,
CC   as verified by in vitro systems.
CC   See also AAQ25452-25501 and AAQ30226-448.
XX
SQ   Sequence 18 BP; 13 A; 0 C; 0 G; 5 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 13;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    159 aaaatataaataaa 172
          ||||||||||||||
Db      2 aaaatataaataaa 15


RESULT  17
AAZ22162/c
ID   AAZ22162 standard; DNA; 18 BP.
XX
AC   AAZ22162;
XX
DT   26-NOV-1999  (first entry)
XX
DE   Human c-IAP-1 mRNA inhibiting antisense oligo ISIS #23344.
XX
KW   Cellular Inhibitor of Apoptosis-1; antisense; diagnostic; therapeutic;
KW   c-IAP-1; prophylaxis; infection; inflammation; tumor formation; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   US5958772-A.
XX
PD   28-SEP-1999.
XX
PF   03-DEC-1998;   98US-0205204.
XX
PR   03-DEC-1998;   98US-0205204.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Cowsert LM,  Ackermann EJ;
XX
DR   WPI; 1999-561047/47.
XX
PT   Antisense compounds complementary to Cellular Inhibitor of Apoptosis-1
PT   useful for e.g. diagnostics, therapeutics, and as research reagents -
XX
PS   Example 15; Column 38; 32pp; English.
XX
CC   The invention provides antisense compounds of 8-30 nucleotides that
CC   inhibit the expression of human Cellular Inhibitor of Apoptosis-1
CC   (c-IAP-1). The antisense compounds may be used for diagnostics,
CC   therapeutics (for modulating the expression of c-IAP-1), prophylaxis
CC   (e.g. to prevent or delay infection, inflammation, or tumor formation),
CC   as research reagents (e.g. to distinguish between members of a
CC   biological pathway) and in kits. Sequences AAZ22150-189 represent
CC   phosphorothioate oligonucleotides used for antisense inhibition of
CC   cellular inhibitor of apoptosis-1.
XX
SQ   Sequence 18 BP; 3 A; 2 C; 1 G; 12 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 20;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    160 aaatataaataaaa 173
          ||||||||||||||
Db     18 AAATATAAATAAAA 5


RESULT  18
AAS01590
ID   AAS01590 standard; DNA; 22 BP.
XX
AC   AAS01590;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human IQGAP2 CpG island 5'-bisulfite PCR primer.
XX
KW   Human; T-type calcium channel; CACNA1G; cytosine methylation; CpG island;
KW   cellular proliferative disorder; colorectal cancer; age related disease;
KW   apolipoprotein B; APOB; caudal type homeobox transcription factor 2;
KW   CDX2; epidermal growth factor receptor; EGFR; fibrillin-1; FBN1;
KW   G protein-coupled receptor 37; GPR37; heat shock 70kD protein 6; HSP70B';
KW   HSPA6; RasGAP-related protein; IQGAP2; proteinase-activated receptor 2;
KW   PAR2; paired-like homeodomain transcription factor 2; PITX2; klotho; KL;
KW   patched A; patched B; PTCHA; PTCHB; syndecan 1; syndecan 4; SDC1; SDC4;
KW   chromosome 5q; PCR primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200119845-A1.
```

```
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-US25479.
XX
PR   15-SEP-1999;   99US-0398522.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Issa J;
XX
DR   WPI; 2001-244777/25.
XX
PT   New nucleic acid molecule for use as a marker for screening cancer,
PT   comprises the coding region for a T-type calcium channel and regulatory
PT   sequences associated with the channel  -
XX
PS   Claim 21; Page 34; 125pp; English.
XX
CC   The present sequence for 5'-bisulfite PCR primer is used to study the
CC   methylation state of human RasGAP-related protein (IQGAP2) which
CC   maps to chromosome 5q. The methylation state of specific regions
CC   within CpG islands associated with a novel T-type calcium channel
CC   CACNA1G gene correlate with several cancerous phenotypes involving
CC   various tissue and cell types. Since aberrant methylation of normally
CC   unmethylated CpG islands is often observed in immortalised and
CC   transformed cells, CACNA1G is implicated in cellular proliferative
CC   disorders e.g. leukaemia, colorectal, lung, breast and other cancers. The
CC   nucleic acid coding for CACNA1G is useful as a marker for screening
CC   cancer and age related diseases. A diagnostic kit containing primers
CC   (AAS01574-AAS01623) for amplification of a CpG-containing nucleic acid,
CC   where the primer hybridises with a target polynucleotide sequence
CC   (AAS01627-AAS01676), can be used for detecting aberrant methylation. The
CC   CpG island sequences (AAS01677-AAS01692) are selected from genes encoding
CC   CACNA1G, apolipoprotein B (APOB), caudal type homeobox transcription
CC   factor 2 (CDX2), epidermal growth factor receptor (EGFR), fibrillin-1
CC   (FBN1), G protein-coupled receptor 37 (GPR37), heat shock 70kD protein 6
CC   (HSP70B'; HSPA6), RasGAP-related protein (IQGAP2), klotho (KL),
CC   proteinase-activated receptor 2 (PAR2), paired-like homeodomain
CC   transcription factor 2 (PITX2), patched A and B (PTCHA; PTCHB) and
CC   syndecan 1 and 4 (SDC1; SDC4) or a MINT31 sequence.
XX
SQ   Sequence 22 BP; 4 A; 0 C; 3 G; 15 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 22;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 6.8e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttta 209
          ||||||||||||||
Db      2 ttatttttatttta 15


RESULT  19
AAS01643/c
ID   AAS01643 standard; DNA; 22 BP.
XX
AC   AAS01643;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human IQGAP2 5'-target sequence for bisulfite PCR.
XX
KW   Human; T-type calcium channel; CACNA1G; cytosine methylation; CpG island;
KW   cellular proliferative disorder; colorectal cancer; age related disease;
KW   apolipoprotein B; APOB; caudal type homeobox transcription factor 2;
KW   CDX2; epidermal growth factor receptor; EGFR; fibrillin-1; FBN1;
KW   G protein-coupled receptor 37; GPR37; heat shock 70kD protein 6; HSP70B';
KW   HSPA6; RasGAP-related protein; IQGAP2; proteinase-activated receptor 2;
KW   PAR2; paired-like homeodomain transcription factor 2; PITX2; klotho; KL;
KW   patched A; patched B; PTCHA; PTCHB; syndecan 1; syndecan 4; SDC1; SDC4;
KW   chromosome 5q; ds.
XX
OS   Homo sapiens.
XX
PN   WO200119845-A1.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-US25479.
XX
PR   15-SEP-1999;   99US-0398522.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Issa J;
XX
DR   WPI; 2001-244777/25.
XX
PT   New nucleic acid molecule for use as a marker for screening cancer,
PT   comprises the coding region for a T-type calcium channel and regulatory
PT   sequences associated with the channel  -
XX
PS   Claim 20; Page 37; 125pp; English.
XX
CC   The present sequence for human RasGAP-related protein (IQGAP2)
CC   5'-target sequence (complementary to the 5'-bisulfite PCR primer) is used
CC   to study the methylation state of IQGAP2 which maps to chromosome 5q. The
CC   methylation state of specific regions within CpG islands associated with
CC   a novel T-type calcium channel CACNA1G gene correlate with several
CC   cancerous phenotypes involving various tissue and cell types. Since
CC   aberrant methylation of normally unmethylated CpG islands is often
CC   observed in immortalised and transformed cells, CACNA1G is implicated in
CC   cellular proliferative disorders e.g. leukaemia, colorectal, lung, breast
CC   and other cancers. The nucleic acid coding for CACNA1G is useful as a
CC   marker for screening cancer and age related diseases. A diagnostic kit
CC   containing primers (AAS01574-AAS01623) for amplification of a
CC   CpG-containing nucleic acid, where the primer hybridises with a target
CC   polynucleotide sequence (AAS01627-AAS01676), can be used for detecting
CC   aberrant methylation. The CpG island sequences (AAS01677-AAS01692) are
CC   selected from genes encoding CACNA1G, apolipoprotein B (APOB), caudal
CC   type homeobox transcription factor 2 (CDX2), epidermal growth factor
CC   receptor (EGFR), fibrillin-1 (FBN1), G protein-coupled receptor 37
CC   (GPR37), heat shock 70kD protein 6 (HSP70B'; HSPA6), IQGAP2, klotho (KL),
CC   proteinase-activated receptor 2 (PAR2), paired-like homeodomain
CC   transcription factor 2 (PITX2), patched A and B (PTCHA; PTCHB) and
CC   syndecan 1 and 4 (SDC1; SDC4) or a MINT31 sequence.
XX
SQ   Sequence 22 BP; 15 A; 3 C; 0 G; 4 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 22;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 6.8e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttta 209
          ||||||||||||||
Db     21 TTATTTTTATTTTA 8


RESULT  20
AAV07952/c
ID   AAV07952 standard; DNA; 26 BP.
XX
AC   AAV07952;
XX
DT   02-FEB-1999  (first entry)
XX
DE   Helicobacter pylori polypeptide GHPO 1414 5' DNA primer.
XX
KW   GHPO 1414; infection; gastritis; ulcer; vaccine; diagnosis;
KW   therapy; PCR; primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
```

XX
PN WO9843479-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06421.
XX
PR 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX
PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J;
XX
DR WPI; 1998-568251/48.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastroduodenal diseases
XX
PS Claim 5; Page 145; 184pp; English.
XX
CC This 5' primer is used with a 3' primer (see AAV07954) in the PCR
CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic
CC DNA in order to obtain DNA (see AAV07921) encoding the unprocessed
CC form of a 76 kDa polypeptide (see AAW73032) designated GHPO 1414.
CC The isolated polynucleotide, and encoded polypeptide, can be used
CC to develop vaccines for the treatment and prevention of Helicobacter
CC infections.
XX
SQ Sequence 26 BP; 14 A; 5 C; 4 G; 3 T; 0 other;

Query Match 4.3%; Score 14; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 ttttcatgttttct 320
||||||||||||||
Db 18 TTTTCATGTTTTCT 5

RESULT 21
AAV07922/c
ID AAV07922 standard; DNA; 26 BP.
XX
AC AAV07922;
XX
DT 02-FEB-1999 (first entry)
XX
DE Helicobacter pylori polypeptide GHPO 386 5' DNA primer.
XX
KW GHPO 386; infection; gastritis; ulcer; vaccine; diagnosis; therapy;
KW PCR; primer; ss.
XX
OS Synthetic.
OS Helicobacter pylori.
XX
PN WO9843479-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06421.
XX
PR 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX
PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J;
XX
DR WPI; 1998-568251/48.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastroduodenal diseases
XX
PS Claim 5; Page 137; 184pp; English.
XX
CC This 5' primer is used with a 3' primer (see AAV07924) in the PCR
CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic
CC DNA in order to obtain DNA (see AAV72001
CC) encoding the unprocessed
CC form of a 76 kDa polypeptide (see AAW73022) designated GHPO 386. The
CC isolated polynucleotide, and encoded polypeptide, can be used to
CC develop vaccines for the treatment and prevention of Helicobacter
CC infections.
XX
SQ Sequence 26 BP; 15 A; 5 C; 4 G; 2 T; 0 other;

Query Match 4.3%; Score 14; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 ttttcatgttttct 320
||||||||||||||
Db 20 TTTTCATGTTTTCT 7

RESULT 22
AAV07937/c
ID AAV07937 standard; DNA; 27 BP.
XX
AC AAV07937;
XX
DT 02-FEB-1999 (first entry)
XX
DE Helicobacter pylori polypeptide GHPO 896 5' DNA primer.
XX
KW GHPO 896; infection; gastritis; ulcer; vaccine; diagnosis;
KW therapy; PCR; primer; ss.
XX
OS Synthetic.
OS Helicobacter pylori.
XX
PN WO9843479-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06421.
XX
PR 01-APR-1997; 97US-0834666.
PR 01-APR-1997; 97US-0831310.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.
XX
PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J;
XX
DR WPI; 1998-568251/48.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastroduodenal diseases
XX
PS Claim 5; Page 141; 184pp; English.
XX
CC This 5' primer is used with a 3' primer (see AAV07939) in the PCR
CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic
CC DNA in order to obtain DNA (see AAV07916) encoding the unprocessed
CC form of a 76 kDa polypeptide (see AAW73027) designated GHPO 896.
CC The isolated polynucleotide, and encoded polypeptide, can be used
CC to develop vaccines for the treatment and prevention of Helicobacter

```
CC   infections.
XX
SQ   Sequence 27 BP; 14 A; 5 C; 4 G; 4 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 19;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 6.7e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    307 ttttcatgttttct 320
          ||||||||||||||
Db     18 TTTTCATGTTTTCT 5


RESULT  23
AAT42655/c
ID   AAT42655 standard; DNA; 29 BP.
XX
AC   AAT42655;
XX
DT   25-FEB-1997  (first entry)
XX
DE   Primer for amplifying verotoxin (VT-1) subunit A coding sequence.
XX
KW   Verotoxin; Escherichia coli; enteric infection; diarrhoea; vaccine;
KW   haemolytic uraemic syndrome; detection; ss.
XX
OS   Synthetic.
XX
PN   WO9630043-A1.
XX
PD   03-OCT-1996.
XX
PF   25-MAR-1996;   96WO-US04093.
XX
PR   24-MAR-1995;   95US-0410058.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Carroll SB,  Padhye NV,  Stafford DC;
XX
DR   WPI; 1996-505779/50.
XX
PT   Compsn. contg. neutralising antitoxin against E.coli vero-toxin -
PT   used to treat intoxicated individuals, and as a prophylactic against
PT   diarrhoeal disease or extra-intestinal complications of E.coli
PT   infection
XX
PS   Example 6; Page 58; 101pp; English.
XX
CC   Compositions containing neutralising antitoxin against one or more E.
CC   coli verotoxin (VT) can be used to treat intoxicated adults and
CC   children with enteric bacterial infections. They may also be used as
CC   prophylactics e.g. as a vaccine, against diarrhoeal disease or the
CC   development of extra-intestinal complications of E.coli infection,
CC   especially haemolytic uraemic syndrome.  The antitoxin can also be
CC   used to detect E. coli VT in a sample. The VT is recombinant,
CC   preferably a fusion protein containing a non-VT protein sequence and
CC   part of the E.coli VT1 or VT2 sequence. Two primers (AAT42655,
CC   AAT42656) were used to amplify the verotoxin VT-1 A subunit coding
CC   sequence and add a histidine tag coding sequence to the subunit
CC   sequence. Two primers (AAT42655, AAT42658) were used to amplify the
CC   verotoxin VT-1 A and B subunits and add a histidine tag coding
CC   sequence to the subunit sequences.
XX
SQ   Sequence 29 BP; 11 A; 2 C; 5 G; 11 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 17;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     22 AAAATAATTATTTT 9


RESULT  24
AAA51200/c
ID   AAA51200 standard; DNA; 29 BP.
XX
AC   AAA51200;
XX
DT   26-SEP-2000  (first entry)
XX
DE   N-terminal primer for E. coli verotoxin 1 subunit A gene.
XX
KW   VT-1; verotoxin; antitoxin therapy; fusion protein; affinity tag; food;
KW   recombinant production; screening; dairy; anti-bacterial; vaccine;
KW   primer; polyhistidine; ss.
XX
OS   Escherichia coli.
OS   Synthetic.
XX
PN   US6080400-A.
XX
PD   27-JUN-2000.
XX
PF   13-MAR-1997;   97US-0816977.
XX
PR   24-MAR-1995;   95US-0410058.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Byrne LM;
XX
DR   WPI; 2000-451195/39.
XX
PT   Bacterial cell for recombinantly expressing bacterial toxins in large
PT   quantities useful for immunization and treatment of bacterial
PT   infections, comprises expression vector encoding bacterial toxin
XX
PS   Example 6; Column 83; 83pp; English.
XX
CC   E. coli verotoxin (VT) type 1 and 2 subunits A and B were cloned into
CC   pET-23b, designed to allow expression of the native proteins containing
CC   C-terminal polyhistidine tags. The VT-1 and VT-2 genes were engineered
CC   to convert the signal sequence methionine codon into a NdeI site to
CC   allow cloning of the amplified genes into the vector without addition of
CC   vector-encoded amino acids. The C-terminal primers comprises the
CC   C-terminal 7 codons of each gene fused to the sequence CTCGAGCC, in order
CC   to add the polyhistidine tag. The primers delete the native stop codons,
CC   and when cloned into pET-23 add a C-terminal extension of Leu-Glu-(His)6.
CC   VT B chains are small proteins (approximately 8 kDa), so use of a small
CC   affinity tag was preferred (i.e. polyhistidine). A polyhistidine affinity
CC   tag facilitates single step affinity purification of subunits from
CC   periplasmic extracts. However, due to poor recovery of his-tagged VT-1 A
CC   and VT-2 A chains, expression of maltose binding protein (MBP) fused
CC   subunits was undertaken. Due to the toxicity of the VT-2 B subunit,
CC   strict uninduced promoter control is necessary to permit cell viability.
CC   Bacterial host cells expressing a recombinant expression vector encoding
CC   a polyhistidine affinity tag and a portion of the VT-2 B chain are
CC   claimed. The vector is chosen from pET24hisVT2BL+, pET24hisVT2BL- and
CC   pET24VT2B, where "L+" indicates that the vector encodes the preprotein
CC   form of the protein and "L-" indicates that the vector encodes the mature
CC   form of the protein. The bacterial cell is capable of expressing large
CC   quantities (40 mg/l) of VT-2B. The toxins are useful for immunizing
CC   non-mammals and for detecting bacterial toxins in environmental samples
CC   including soil, water, industrial samples, biological samples and samples
CC   obtained from food and dairy processing instruments.
XX
SQ   Sequence 29 BP; 11 A; 2 C; 5 G; 11 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 21;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+03;
```

```
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     22 AAAATAATTATTTT 9


RESULT  25
AAT25703/c
ID   AAT25703 standard; cDNA to mRNA; 31 BP.
XX
AC   AAT25703;
XX
DT   10-OCT-1996  (first entry)
XX
DE   Human gene signature HUMGS07904.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 1910; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 31 BP; 14 A; 3 C; 3 G; 10 T; 1 other;


  Query Match             4.3%;  Score 14;  DB 16;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    198 atttttattttaaa 211
          ||||||||||||||
Db     20 ATTTTTATTTTAAA 7


RESULT  26
AAV67854
ID   AAV67854 standard; DNA; 31 BP.
XX
AC   AAV67854;
XX
DT   24-DEC-1998  (first entry)
XX
DE   Nucleotide fragment containing polymorphic site, WI-11163.
XX
KW   ss; polymorphic site; nucleic acid analysis; diagnosis; monitoring;
KW   cancer; inflammation; heart disease; CNS disease.
XX
OS   Homo sapiens.
XX
PN   WO9838846-A2.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;   98WO-US04571.
XX
PR   28-MAR-1997;   97US-0042125.
PR   07-MAR-1997;   97US-0813159.
XX
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Berno A,  Chee M,  Fan J,  Lipshutz RJ;
XX
DR   WPI; 1998-495419/42.
XX
PT   New nucleic acid segments containing polymorphic sites, or
PT   complements and methods of detecting a nucleic acid - for general
PT   use including diagnosis and monitoring of diseases
XX
PS   Claim 1; Page 25; 42pp; English.
XX
CC   New nucleic acid segment comprising one of the 10 - 100 bp sequences
CC   given in the specification (sequences of a polymorphic site), or the
CC   complement of the segment and a method of analysing a nucleic acid
CC   comprising determining the base occupying the polymorphic site of the
CC   polymorphic fragment sequences are disclosed in the specification.  The
CC   information obtained from nucleic acid analysis by the method described
CC   is useful in diagnosis or monitoring of diseases like cancer,
CC   inflammation, heart disease, CNS diseases, and susceptibility to
CC   infection by microorganisms. In addition, the nucleic acid segments are
CC   useful in manufacturing medication in the treatment of prophylaxis of
CC   diseases, and also the use of the DNA segments as pharmaceutical.
XX
SQ   Sequence 31 BP; 17 A; 2 C; 4 G; 7 T; 1 other;


  Query Match             4.3%;  Score 14;  DB 19;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    209 aaatttgttaaaaa 222
          ||||||||||||||
Db     17 aaatttgttaaaaa 30


RESULT  27
AAZ27689/c
ID   AAZ27689 standard; DNA; 32 BP.
XX
AC   AAZ27689;
XX
DT   22-DEC-1999  (first entry)
XX
DE   PCR primer for Verotoxin gene.
XX
KW   Verotoxin; VT1; VT2; detection; PCR primer; ss.
XX
OS   Synthetic.
```

```
OS   Escherichia coli.
XX
PN   JP11243996-A.
XX
PD   14-SEP-1999.
XX
PF   27-FEB-1998;   98JP-0047677.
XX
PR   27-FEB-1998;   98JP-0047677.
XX
PA   (TOYM ) TOYOBO KK.
XX
DR   WPI; 1999-603716/52.
XX
PT   An oligonucleotide for amplification of verotoxin - useful in the
PT   detection of inactivated verotoxin gene by transfer of a foreign DNA
PT   fragment
XX
PS   Claim 11; Page 9; 10pp; Japanese.
XX
CC   This sequence represents a PCR primer of the invention. The primer is
CC   used for amplification of the E. coli verotoxin (VT) gene. The
CC   oligonucleotide is useful for detection of inactivated VT gene by
CC   transfer of a foreign DNA fragment. Simple, rapid and specific
CC   amplification of VT gene from environmental factors is achieved using the
CC   oligonucleotide of the invention.
XX
SQ   Sequence 32 BP; 12 A; 2 C; 4 G; 14 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 20;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     17 AAAATAATTATTTT 4


RESULT  28
AAT97603/c
ID   AAT97603 standard; DNA; 36 BP.
XX
AC   AAT97603;
XX
DT   30-APR-1998  (first entry)
XX
DE   Shigella dysenteriae delta-stxA allele PCR primer 13.
XX
KW   Delta-virG allele; delta-guaB-A allele; PCR; amplification; primer;
KW   delta-stxA allele; shigellosis; vaccine; ss.
XX
OS   Synthetic.
OS   Shigella dysenteriae.
XX
PN   WO9737685-A1.
XX
PD   16-OCT-1997.
XX
PF   09-APR-1997;   97WO-US05954.
XX
PR   09-APR-1996;   96US-0629600.
XX
PA   (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI   Levine MM,  Noriega FR;
XX
DR   WPI; 1997-512417/47.
XX
PT   Shigella mutants with mutation in guaB-A - used in vaccines against
PT   Shigellosis
XX
PS   Example 6; Page 57; 94pp; English.
XX
CC   This is a PCR primer used in the amplification of the Shigella
CC   dysenteriae 1 delta-stxA allele.  The delta-stxA allele was integrated
CC   into delta-guaB-A of delta-guaB-A, delta-virG S. dysenteriae 1, which
CC   inactivated the shiga toxin of this strain.  The mutant can be used in
CC   the preparation of vaccines such as, a live vector vaccine comprising
CC   a Shigella mutant, (which encodes and expresses a foreign antigen, and
CC   a pharmaceutically acceptable carrier) or a DNA mediated vaccine
CC   comprising the Shigella mutant (which also contains a plasmid which
CC   encodes and expresses a foreign antigen in a eukaryotic cell).  The
CC   vaccines can be used against Shigellosis.
XX
SQ   Sequence 36 BP; 11 A; 3 C; 10 G; 12 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 18;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 aaaataattatttt 202
          ||||||||||||||
Db     27 AAAATAATTATTTT 14


RESULT  29
AAC90606
ID   AAC90606 standard; RNA; 36 BP.
XX
AC   AAC90606;
XX
DT   20-MAR-2001  (first entry)
XX
DE   Tomato spotted wilt virus S RNA partial sequence #10.
XX
KW   Tospovirus resistance; transgenic plant; tomato spotted wilt virus;
KW   Impatiens necrotic spot virus; TSWV; ss.
XX
OS   Tomato spotted wilt virus.
XX
FH   Key             Location/Qualifiers
FT   misc_binding     1..13
FT                    /*tag=  a
FT                    /bound_moiety= "binds nucleotides 33-21 of AAC89655"
FT   misc_binding     15..28
FT                    /*tag=  b
FT                    /bound_moiety= "binds nucleotides 19-6 of AAC89655"
FT   misc_binding     32..36
FT                    /*tag=  c
FT                    /bound_moiety= "binds nucleotides 5-1 of AAC89655"
XX
PN   US6150585-A.
XX
PD   21-NOV-2000.
XX
PF   26-NOV-1996;   96US-0757011.
XX
PR   02-MAY-1991;   91US-0694734.
PR   14-APR-1993;   93US-0047346.
PR   26-OCT-1993;   93US-0143397.
PR   27-JUL-1994;   94US-0280903.
PR   03-NOV-1989;   89US-0431259.
PR   05-DEC-1989;   89US-0446024.
XX
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Peters D,  Gielen JJL,  De Haan PT,  Van Grinsven MQJM,  Kool AJ;
PI   Goldbach RW;
XX
DR   WPI; 2001-060031/07.
XX
PT   Recombinant DNA construct comprising a DNA sequence encoding an RNA
PT   sequence that codes for a tospovirus protein, useful for producing
PT   plants with reduced susceptibility to tospovirus infection -
```

```
XX
PS   Example 9; Fig 16C; 49pp; English.
XX
CC   The present invention provides DNA constructs encoding RNA sequences from
CC   a tospovirus which can be used to produce transgenic plants with immunity
CC   to tospoviruses. Examples of tospoviruses include the tomato spotted wilt
CC   virus and the Impatiens necrotic spot virus.
XX
SQ   Sequence 36 BP; 7 A; 0 C; 0 G; 29 U; 0 other;


  Query Match             4.3%;  Score 14;  DB 22;  Length 36;
  Best Local Similarity   21.4%;  Pred. No. 6.5e+03;
  Matches    3;  Conservative   11;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    196 ttatttttatttta 209
          ::|:::::|::::|
Db      7 uuauuuuuauuuua 20


RESULT  30
AAQ62952
ID   AAQ62952 standard; DNA; 37 BP.
XX
AC   AAQ62952;
XX
DT   09-SEP-1994  (first entry)
XX
DE   Glycophorin antibody 1C3 Fab coding region PCR primer.
XX
KW   Glycophorin; antibody 1C3; target binding polypeptide; PCR;
KW   polymerase chain reaction; primer; antibody engineering;
KW   humanized antibody; phagemid pHFA; plasmid p569;ss.
XX
OS   Synthetic.
XX
PN   WO9407921-A.
XX
PD   14-APR-1994.
XX
PF   24-SEP-1993;   93WO-AU00491.
XX
PR   25-SEP-1992;   92AU-0004973.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Atwell JL,  Colman PM,  Hudson PJ,  Irving RA,  Kortt A;
PI   Lah M,  Malbyrl,  Power BE;
XX
DR   WPI; 1994-135515/16.
XX
PT   New target-binding polypeptide(s) used for diagnosis, etc. -
PT   having a stable core polypeptide region with at least one
PT   target-binding region covalently attached, opt. mutated to alter
PT   specificity, etc.
XX
PS   Example; Page 36; 67pp; English.
XX
CC   PCR primers given in AAQ62951-52 were used to clone anti-glycophorin
CC   antibody 1C3 Fab coding region.  The DNA sequence of the first 1443
CC   bases of the Fab fragment in pHFA, ready for ligation post PCR
CC   amplification for ligation into p569, is given in AAQ62958.
XX
SQ   Sequence 37 BP; 14 A; 5 C; 0 G; 18 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 15;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     83 aaatttataattat 96
          ||||||||||||||
Db      1 aaatttataattat 14


RESULT  31
AAF55449/c
ID   AAF55449 standard; DNA; 45 BP.
XX
AC   AAF55449;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Oligonucleotide used to construct the shuttle vector pAAO-E-TATA.
XX
KW   Adenovirus vector; gene delivery vector; E2B gene; E4 gene; vaccine;
KW   replication defective virus; cell proliferation; cell differentiation;
KW   gene therapy; ss.
XX
OS   Synthetic.
XX
PN   EP1083229-A1.
XX
PD   14-MAR-2001.
XX
PF   10-SEP-1999;   99EP-0202966.
XX
PR   10-SEP-1999;   99EP-0202966.
XX
PA   (INTR-) INTROGENE BV.
XX
DR   WPI; 2001-228258/24.
XX
PT   Producing a recombinant adenovirus-like gene delivery vehicle with
PT   modified E2B or E4 functions, for gene therapy, comprises generating an
PT   adenoviral vector where E2B or E4 is under the control of a synthetic
PT   promoter -
XX
PS   Example 1; Page 10; 56pp; English.
XX
CC   The specification describes a method for producing a recombinant
CC   adenovirus-like gene delivery vehicle having reduced expression of
CC   adenoviral E2B and/or E4 gene products in a target cell. The method
CC   comprises generating a recombinant adenoviral vector lacking E1A and
CC   E1B sequences, but having at least the E2B and/or E4 sequences encoding
CC   products essential for adenoviral replication. Compared with previous
CC   vectors, the new recombinant vectors are replication defective and
CC   express the remaining viral genes only at background levels. The vector
CC   itself does not dominantly elicit a response of the immune system, but
CC   the immune response is directed primarily against the transgene product,
CC   and is less toxic to the cells which, in turn, results in a prolonged
CC   synthesis of the protein of interest. The recombinant adenoviral
CC   vectors are used in the functional characterization of gene products in
CC   vitro or in vivo, and for overexpression of known and novel genes in
CC   cells lines, tissues or animals in order to find genes that encode for
CC   proteins with a desired function such as those that interfere with cell
CC   proliferation and differentiation. The vectors are further used as
CC   vaccines, in gene therapy, and for protein production in mammalian
CC   cells. Oligonucleotides AAF55449-50 were annealed together, and used to
CC   construct a shuttle vector which was used in the method of the invention.
XX
SQ   Sequence 45 BP; 14 A; 5 C; 7 G; 19 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 22;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     79 ctgaaaatttataa 92
          ||||||||||||||
Db     21 CTGAAAATTTATAA 8


RESULT  32
AAF55450
ID   AAF55450 standard; DNA; 45 BP.
```

```
XX
AC   AAF55450;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Oligonucleotide used to construct the shuttle vector pAAO-E-TATA.
XX
KW   Adenovirus vector; gene delivery vector; E2B gene; E4 gene; vaccine;
KW   replication defective virus; cell proliferation; cell differentiation;
KW   gene therapy; ss.
XX
OS   Synthetic.
XX
PN   EP1083229-A1.
XX
PD   14-MAR-2001.
XX
PF   10-SEP-1999;   99EP-0202966.
XX
PR   10-SEP-1999;   99EP-0202966.
XX
PA   (INTR-) INTROGENE BV.
XX
DR   WPI; 2001-228258/24.
XX
PT   Producing a recombinant adenovirus-like gene delivery vehicle with
PT   modified E2B or E4 functions, for gene therapy, comprises generating an
PT   adenoviral vector where E2B or E4 is under the control of a synthetic
PT   promoter  -
XX
PS   Example 1; Page 10; 56pp; English.
XX
CC   The specification describes a method for producing a recombinant
CC   adenovirus-like gene delivery vehicle having reduced expression of
CC   adenoviral E2B and/or E4 gene products in a target cell. The method
CC   comprises generating a recombinant adenoviral vector lacking E1A and
CC   E1B sequences, but having at least the E2B and/or E4 sequences encoding
CC   products essential for adenoviral replication. Compared with previous
CC   vectors, the new recombinant vectors are replication defective and
CC   express the remaining viral genes only at background levels. The vector
CC   itself does not dominantly elicit a response of the immune system, but
CC   the immune response is directed primarily against the transgene product,
CC   and is less toxic to the cells which, in turn, results in a prolonged
CC   synthesis of the protein of interest. The recombinant adenoviral
CC   vectors are used in the functional characterization of gene products in
CC   vitro or in vivo, and for overexpression of known and novel genes in
CC   cells lines, tissues or animals in order to find genes that encode for
CC   proteins with a desired function such as those that interfere with cell
CC   proliferation and differentiation. The vectors are further used as
CC   vaccines, in gene therapy, and for protein production in mammalian
CC   cells. Oligonucleotides AAF55449-50 were annealed together, and used to
CC   construct a shuttle vector which was used in the method of the invention.
XX
SQ   Sequence 45 BP; 19 A; 7 C; 5 G; 14 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 22;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      29 ctgaaaatttataa 42


RESULT  33
AAC88874/c
ID   AAC88874 standard; DNA; 45 BP.
XX
AC   AAC88874;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Oligonucleotide TATAplus.
XX
KW   Adenovirus type 35; Ad35; adenovirus type 5; Ad5; gene delivery vehicle;
KW   gene therapy; linker; ss.
XX
OS   Synthetic.
XX
PN   EP1054064-A1.
XX
PD   22-NOV-2000.
XX
PF   16-MAY-2000; 2000EP-0201738.
XX
PR   17-MAY-1999;   99EP-0201545.
XX
PA   (INTR-) INTROGENE BV.
XX
PI   Bout A,  Vogels R,  Havenga MJE;
XX
DR   WPI; 2001-001097/01.
XX
PT   Adenovirus derived gene delivery vehicles comprising at least one
PT   element of adenovirus type 35, efficiently transfers genetic material
PT   to a human cell without the inherent problem of pre-existing immunity -
XX
PS   Example 13; Page 31; 138pp; English.
XX
CC   The present sequence is a linker used in a gene delivery vehicle
CC   comprising an element of adenovirus type 35 or a functional equivalent
CC   of such an element. The element is responsible for avoiding or reducing
CC   neutralising activity against adenoviral elements by the host to which
CC   the gene is to be delivered. The vehicle can be used to deliver genes or
CC   nucleic acids of interest to host cells. Use of the delivery system
CC   efficiently transfers genetic material to a human cell without the
CC   inherent problem of pre-existing immunity, found with previous serotypes.
XX
SQ   Sequence 45 BP; 14 A; 5 C; 7 G; 19 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 22;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      21 CTGAAAATTTATAA 8


RESULT  34
AAC88875
ID   AAC88875 standard; DNA; 45 BP.
XX
AC   AAC88875;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Oligonucleotide TATAmin.
XX
KW   Adenovirus type 35; Ad35; adenovirus type 5; Ad5; gene delivery vehicle;
KW   gene therapy; linker; ss.
XX
OS   Synthetic.
XX
PN   EP1054064-A1.
XX
PD   22-NOV-2000.
XX
PF   16-MAY-2000; 2000EP-0201738.
XX
PR   17-MAY-1999;   99EP-0201545.
XX
PA   (INTR-) INTROGENE BV.
XX
```

```
PI   Bout A,  Vogels R,  Havenga MJE;
XX
DR   WPI; 2001-001097/01.
XX
PT   Adenovirus derived gene delivery vehicles comprising at least one
PT   element of adenovirus type 35, efficiently transfers genetic material
PT   to a human cell without the inherent problem of pre-existing immunity -
XX
PS   Example 13; Page 31; 138pp; English.
XX
CC   The present sequence is a linker used in a gene delivery vehicle
CC   comprising an element of adenovirus type 35 or a functional equivalent
CC   of such an element. The element is responsible for avoiding or reducing
CC   neutralising activity against adenoviral elements by the host to which
CC   the gene is to be delivered. The vehicle can be used to deliver genes or
CC   nucleic acids of interest to host cells. Use of the delivery system
CC   efficiently transfers genetic material to a human cell without the
CC   inherent problem of pre-existing immunity, found with previous serotypes.
XX
SQ   Sequence 45 BP; 19 A; 7 C; 5 G; 14 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 22;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 ctgaaaatttataa 92
           ||||||||||||||
Db      29 ctgaaaatttataa 42


RESULT  35
AAZ66366
ID   AAZ66366 standard; DNA; 47 BP.
XX
AC   AAZ66366;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:713.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,C)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 1; Page 382; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 22 A; 2 C; 6 G; 17 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 21;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     109 gttaataaataaaa 122
           ||||||||||||||
Db       5 gttaataaataaaa 18


RESULT  36
AAZ67473/c
ID   AAZ67473 standard; DNA; 47 BP.
XX
AC   AAZ67473;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:1820.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,C)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 1; Page 618; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
```

```
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 16 A; 4 C; 2 G; 25 T; 0 other;


  Query Match          4.3%;  Score 14;  DB 21;  Length 47;
  Best Local Similarity  100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     162 atataaataaaatt 175
           ||||||||||||||
Db      23 ATATAAATAAAATT 10


RESULT  37
AAZ67533/c
ID   AAZ67533 standard; DNA; 47 BP.
XX
AC   AAZ67533;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:1880.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,C)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 1; Page 631; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 15 A; 6 C; 6 G; 20 T; 0 other;


  Query Match          4.3%;  Score 14;  DB 21;  Length 47;
  Best Local Similarity  100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     95 atatgataagtata 108
          ||||||||||||||
Db     30 ATATGATAAGTATA 17


RESULT  38
AAZ67549/c
ID   AAZ67549 standard; DNA; 47 BP.
XX
AC   AAZ67549;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:1896.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,G)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 1; Page 635; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
```

```
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 17 A; 10 C; 2 G; 18 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 21;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    148 ttttaaaaaggaaa 161
          ||||||||||||||
Db     27 TTTTAAAAAGGAAA 14


RESULT  39
AAZ67813
ID   AAZ67813 standard; DNA; 47 BP.
XX
AC   AAZ67813;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:2160.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,G)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9954500-A2.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1999;   99WO-IB00822.
XX
PR   21-APR-1998;   98US-0082614.
PR   23-NOV-1998;   98US-0109732.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I;
XX
DR   WPI; 2000-013267/01.
XX
PT   Novel biallelic markers used to construct a high density disequilibrium
PT   map of the human genome  -
XX
PS   Claim 1; Page 691; 2745pp; English.
XX
CC   AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC   invention, which contain a polymorphic base at position 24 of their
CC   nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC   primers for the biallelic markers. The biallelic markers of the
CC   invention have a variety of uses: they can be used for high density
CC   mapping of the human genome, and in complex association studies and
CC   haplotyping studies which are useful in determining the genetic basis
CC   for disease states. Compositions and methods of the invention can also
CC   be useful for the identification of the targets for the development of
CC   pharmaceutical agents and diagnostic methods, as well as the
CC   characterisation of the differential efficacious responses to and side
CC   effects from pharmaceutical agents acting on a disease as well as other
CC   treatment.
CC   N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC   and 3367, are not actually given a sequence in the Sequence Listing
CC   from the present invention.
XX
SQ   Sequence 47 BP; 19 A; 4 C; 8 G; 16 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 21;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    198 atttttatttttaaa 211
          ||||||||||||||
Db     11 atttttatttttaaa 24


RESULT  40
AAA98312
ID   AAA98312 standard; DNA; 50 BP.
XX
AC   AAA98312;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Human MSH6 fragment 8/exon 8 to 10 DNA Ref-Seq fragment.
XX
KW   Human mismatch repair gene; hMSH6; disease predisposition; genotype;
KW   mutation; carcinoma; colorectal; endometrial; ovarian; leukemia;
KW   neoplastic disease; drug development; ds.
XX
OS   Homo sapiens.
XX
PN   DE19909878-A1.
XX
PD   07-SEP-2000.
XX
PF   06-MAR-1999;   99DE-1009878.
XX
PR   06-MAR-1999;   99DE-1009878.
XX
PA   (UYDR ) UNIV DRESDEN TECH.
XX
PI   Plaschke J,  Kruppa C,  Schackert H;
XX
DR   WPI; 2000-588378/56.
XX
PT   Novel variants of the human mismatch repair gene, MSH6, useful e.g. for
PT   determining predisposition to cancer and for development of drugs  -
XX
PS   Disclosure; Fig 3; 14pp; German.
XX
CC   This invention describes a novel method of determining a predisposition
CC   to disease by genotyping a subject's DNA sequence (A) of the human
CC   mismatch repair gene, MSH6 at specified positions and comparing with
CC   reference DNA sequences, optionally taking into account all possible
CC   combinations of variations of the individual mutations, including any
CC   chosen absolute number of variations. (A), and analysis of their
CC   sequences, are useful for the following: (i) determining a predisposition
CC   to disease, especially colorectal, endometrial and ovarian carcinoma and
CC   leukemia; (ii) determining an increased mutation rate (frequency of base
CC   substitutions, insertions and/or deletions) in eukaryotic cells; (iii)
CC   predicting the progression, severity and survival time of patients with
CC   neoplastic disease; (iv) the development of therapeutic and 'life-style'
CC   drugs; (v) predicting individual differences in response to known
CC   chemotherapeutic agents (e.g. cis-platin) or drugs developed from (iv);
CC   (vi) optimizing individual treatments and interventions against
CC   neoplasia; (vii) controlling the mutation rate in eukaryotic cells, in
CC   vitro or in vivo; (viii) constructing genes and vectors, particularly for
```

```
CC   development of pharmaceuticals; (ix) developing diagnostic kits and other
CC   systems for genotyping; and (x) developing in vivo and in vitro test
CC   systems for expressing individual forms of the MSH6 gene, e.g. for
CC   studying pathophysiology of disease or processes in which MSH6 is
CC   involved, and for drug development and testing.
XX
SQ   Sequence 50 BP; 8 A; 8 C; 5 G; 29 T; 0 other;


  Query Match             4.3%;  Score 14;  DB 21;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 6.3e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    140 ctttttgttttaa 153
          ||||||||||||||
Db     11 ctttttgttttaa 24


RESULT  41
AAF48097
ID   AAF48097 standard; DNA; 15 BP.
XX
AC   AAF48097;
XX
DT   30-MAR-2001  (first entry)
XX
DE   IGFBP3 oligonucleotide #1517.
XX
KW   Antisense therapy; antiproliferative; antiinflammatory; antipsoriatic;
KW   cytostatic; dermatological; cardiant; virucide; ophthalmological; keloid;
KW   skin disorder; Insulin-like Growth Factor 1 receptor; IGF-1; pityriasis;
KW   IGF binding protein; IGFBP-2; IGFBP3; inflammation; psoriasis; pilaris;
KW   growth factor mediated cell proliferation; ichthyosis; serborrhoea; ruba;
KW   keratosis; neoplasia; scleroderma; wart; skin cancer; sclerotic disease;
KW   hyperneovascular condition; hyperplasia; kidney disease;
KW   neovascular condition of the retina; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078341-A1.
XX
PD   28-DEC-2000.
XX
PF   21-JUN-2000; 2000WO-AU00693.
XX
PR   21-JUN-1999;   99US-0140345.
XX
PA   (MURD-) MURDOCH CHILDRENS RES INST.
XX
PI   Wraight CJ,  Werther GA,  Edmondson SR;
XX
DR   WPI; 2001-041421/05.
XX
PT   Ameliorating the effects of a disorder, e.g. psoriasis, by
PT   administering UV (ultra-violet) treatment (optional) and an antisense
PT   nucleic acid that inhibits or reduces growth factor mediated cell
PT   proliferation and/or inflammation -
XX
PS   Example 7; Page 54; 201pp; English.
XX
CC   The present invention relates to a method for ameliorating the effects
CC   of skin disorders. The method comprises contacting the skin with an
CC   antisense oligonucleotide, (for Insulin-like Growth Factor [IGF]-1
CC   receptor, IGF binding protein [IGFBP]-2 or IGFBP3), which is capable of
CC   inhibiting or reducing growth factor mediated cell proliferation,
CC   inflammation and/or other disorders. The present sequence is an
CC   oligonucleotide which can be used to design the antisense
CC   oligonucleotides of the present invention (see AAF45151 and
CC   AAF45153-F45161). The method is useful for ameliorating the effects of
CC   psoriasis, ichthyosis, pityriasis, ruba, pilaris, serborrhoea, keloids,
CC   keratosis, neoplasias, scleroderma, warts, benign growths, cancers of the
CC   skin, a hyperneovascular condition such as a neovascular condition of the
CC   retina, brain or skin, growth factor-mediated malignancies, other
CC   sclerotic disease, kidney disease, hyperproliferation of the inside of
CC   blood vessels or any other hyperplasia.
XX
SQ   Sequence 15 BP; 7 A; 1 C; 3 G; 4 T; 0 other;


  Query Match             4.0%;  Score 13;  DB 22;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     26 agaagaatgttta 38
          |||||||||||||
Db      3 agaagaatgttta 15


RESULT  42
AAF48098
ID   AAF48098 standard; DNA; 15 BP.
XX
AC   AAF48098;
XX
DT   30-MAR-2001  (first entry)
XX
DE   IGFBP3 oligonucleotide #1518.
XX
KW   Antisense therapy; antiproliferative; antiinflammatory; antipsoriatic;
KW   cytostatic; dermatological; cardiant; virucide; ophthalmological; keloid;
KW   skin disorder; Insulin-like Growth Factor 1 receptor; IGF-1; pityriasis;
KW   IGF binding protein; IGFBP-2; IGFBP3; inflammation; psoriasis; pilaris;
KW   growth factor mediated cell proliferation; ichthyosis; serborrhoea; ruba;
KW   keratosis; neoplasia; scleroderma; wart; skin cancer; sclerotic disease;
KW   hyperneovascular condition; hyperplasia; kidney disease;
KW   neovascular condition of the retina; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078341-A1.
XX
PD   28-DEC-2000.
XX
PF   21-JUN-2000; 2000WO-AU00693.
XX
PR   21-JUN-1999;   99US-0140345.
XX
PA   (MURD-) MURDOCH CHILDRENS RES INST.
XX
PI   Wraight CJ,  Werther GA,  Edmondson SR;
XX
DR   WPI; 2001-041421/05.
XX
PT   Ameliorating the effects of a disorder, e.g. psoriasis, by
PT   administering UV (ultra-violet) treatment (optional) and an antisense
PT   nucleic acid that inhibits or reduces growth factor mediated cell
PT   proliferation and/or inflammation -
XX
PS   Example 7; Page 54; 201pp; English.
XX
CC   The present invention relates to a method for ameliorating the effects
CC   of skin disorders. The method comprises contacting the skin with an
CC   antisense oligonucleotide, (for Insulin-like Growth Factor [IGF]-1
CC   receptor, IGF binding protein [IGFBP]-2 or IGFBP3), which is capable of
CC   inhibiting or reducing growth factor mediated cell proliferation,
CC   inflammation and/or other disorders. The present sequence is an
CC   oligonucleotide which can be used to design the antisense
CC   oligonucleotides of the present invention (see AAF45151 and
CC   AAF45153-F45161). The method is useful for ameliorating the effects of
CC   psoriasis, ichthyosis, pityriasis, ruba, pilaris, serborrhoea, keloids,
CC   keratosis, neoplasias, scleroderma, warts, benign growths, cancers of the
CC   skin, a hyperneovascular condition such as a neovascular condition of the
CC   retina, brain or skin, growth factor-mediated malignancies, other
CC   sclerotic disease, kidney disease, hyperproliferation of the inside of
CC   blood vessels or any other hyperplasia.
XX
```

```
SQ   Sequence 15 BP; 7 A; 0 C; 3 G; 5 T; 0 other;


  Query Match             4.0%;  Score 13;  DB 22;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 agaagaatgttta 38
           |||||||||||||
Db       2 agaagaatgttta 14


RESULT  43
AAF48099
ID   AAF48099 standard; DNA; 15 BP.
XX
AC   AAF48099;
XX
DT   30-MAR-2001  (first entry)
XX
DE   IGFBP3 oligonucleotide #1519.
XX
KW   Antisense therapy; antiproliferative; antiinflammatory; antipsoriatic;
KW   cytostatic; dermatological; cardiant; virucide; ophthalmological; keloid;
KW   skin disorder; Insulin-like Growth Factor 1 receptor; IGF-1; pityriasis;
KW   IGF binding protein; IGFBP-2; IGFBP3; inflammation; psoriasis; pilaris;
KW   growth factor mediated cell proliferation; ichthyosis; serborrhoea; ruba;
KW   keratosis; neoplasia; scleroderma; wart; skin cancer; sclerotic disease;
KW   hyperneovascular condition; hyperplasia; kidney disease;
KW   neovascular condition of the retina; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078341-A1.
XX
PD   28-DEC-2000.
XX
PF   21-JUN-2000; 2000WO-AU00693.
XX
PR   21-JUN-1999;   99US-0140345.
XX
PA   (MURD-) MURDOCH CHILDRENS RES INST.
XX
PI   Wraight CJ,  Werther GA,  Edmondson SR;
XX
DR   WPI; 2001-041421/05.
XX
PT   Ameliorating the effects of a disorder, e.g. psoriasis, by
PT   administering UV (ultra-violet) treatment (optional) and an antisense
PT   nucleic acid that inhibits or reduces growth factor mediated cell
PT   proliferation and/or inflammation -
XX
PS   Example 7; Page 54; 201pp; English.
XX
CC   The present invention relates to a method for ameliorating the effects
CC   of skin disorders. The method comprises contacting the skin with an
CC   antisense oligonucleotide, (for Insulin-like Growth Factor [IGF]-1
CC   receptor, IGF binding protein [IGFBP]-2 or IGFBP3), which is capable of
CC   inhibiting or reducing growth factor mediated cell proliferation,
CC   inflammation and/or other disorders. The present sequence is an
CC   oligonucleotide which can be used to design the antisense
CC   oligonucleotides of the present invention (see AAF45151 and
CC   AAF45153-F45161). The method is useful for ameliorating the effects of
CC   psoriasis, ichthyosis, pityriasis, ruba, pilaris, serborrhoea, keloids,
CC   keratosis, neoplasias, scleroderma, warts, benign growths, cancers of the
CC   skin, a hyperneovascular condition such as a neovascular condition of the
CC   retina, brain or skin, growth factor-mediated malignancies, other
CC   sclerotic disease, kidney disease, hyperproliferation of the inside of
CC   blood vessels or any other hyperplasia.
XX
SQ   Sequence 15 BP; 6 A; 0 C; 4 G; 5 T; 0 other;


  Query Match             4.0%;  Score 13;  DB 22;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 agaagaatgttta 38
           |||||||||||||
Db       1 agaagaatgttta 13


RESULT  44
AAA57758/c
ID   AAA57758 standard; DNA; 16 BP.
XX
AC   AAA57758;
XX
DT   20-OCT-2000  (first entry)
XX
DE   Nucleotide sequence which is bound by Z2 domain of RIP60 polypeptide.
XX
KW   Human; RIP60; zinc finger protein; nucleic acid delivery complex;
KW   nucleic acid binding domain; nucleic acid condensation domain; ss.
XX
OS   Synthetic.
XX
PN   WO200040723-A2.
XX
PD   13-JUL-2000.
XX
PF   04-JAN-2000; 2000WO-US00212.
XX
PR   04-JAN-1999;   99US-0114743.
PR   04-JAN-1999;   99US-0114745.
XX
PA   (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
XX
PI   Heintz NH,  Houchens CR;
XX
DR   WPI; 2000-465985/40.
XX
PT   Non-viral nucleic acid delivery complex for delivering a nucleic acid
PT   molecule into a cell comprises a modular polypeptide -
XX
PS   Example 17; Page 74; 115pp; English.
XX
CC   AAA57752-66 represent sequences which are bound by the Z2 domain of
CC   the human RIP60 polypeptide. RIP60 is a zinc finger protein. The
CC   nucleic acid binding domain of the RIP60 polypeptide is used to
CC   construct a non-viral nucleic acid delivery complex comprising a
CC   modular polypeptide. The complex comprises a modular peptide containing
CC   a nucleic acid binding domain and a nucleic acid condensation domain
CC   that bind with and condense a nucleic acid molecule of more than
CC   50 kilobases in length. The complex also comprises one or more
CC   polypeptides selected from a cell recognition domain, a protein
CC   transduction domain, a protein degradation domain, an intracellular
CC   targeting domain, a protein interaction domain, an epitope domain and
CC   a protein purification domain. The complexes are used to deliver a
CC   nucleic acid to a cell. The nucleic acids delivered are of various
CC   sizes and preferably greater than 50 kilobases, especially more than
CC   100 or more than 200 kilobases in length.
XX
SQ   Sequence 16 BP; 5 A; 0 C; 0 G; 11 T; 0 other;


  Query Match             4.0%;  Score 13;  DB 21;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     110 ttaataaataaaa 122
           |||||||||||||
Db      15 TTAATAAATAAAA 3


RESULT  45
```

AAQ92084/c
ID AAQ92084 standard; cDNA; 17 BP.
XX
AC AAQ92084;
XX
DT 07-JAN-1996 (first entry)
XX
DE Renilla reniformis luciferase DNA probe-1.
XX
KW Luciferase; enzyme; bioluminescence; luminescence; label; DNA probe;
KW antibody; oligonucleotide; ss.
XX
OS Synthetic.
XX
PN US5418155-A.
XX
PD 23-MAY-1995.
XX
PF 29-DEC-1989; 89US-0458952.
XX
PR 29-DEC-1989; 89US-0458952.
PR 20-AUG-1992; 92US-0933017.
PR 17-JUN-1993; 93US-0079700.
PR 14-DEC-1993; 93US-0167650.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Cormier MJ, Lorenz WW;
XX
DR WPI; 1995-199740/26.
XX
PT New recombinant Renilla luciferase polypeptide - used as a
PT luminescent tag, partic in bio-luminescence assays and for the prodn
PT of antibodies
XX
PS Disclosure; Fig. 4; 18pp; English.
XX
CC This 17-mer oligonucleotide DNA probe, along with Probe-2 (AAQ92085)
CC are used to screen an R. reniformis cDNA library to isolate cDNA
CC encoding Renilla luciferase. The luciferase was then expressed
CC using E. coli.
XX
SQ Sequence 17 BP; 6 A; 0 C; 2 G; 9 T; 0 other;

Query Match 4.0%; Score 13; DB 16; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.8e+04;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    271 aaaaaaattattt 283
          |||||||||||||
Db     15 AAAAAAATTATTT 3
```

Search completed: January 24, 2002, 03:28:19
Job time: 3671 sec

Can't access numbers!?

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:22:17 ; Search time 216.42 Seconds
(without alignments)
356.526 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	90	100.0	1392	20	AAX60299	DNA encoding the b
	2	60	66.7	327	20	AAX60300	Promoter of the be
c	3	32.6	36.2	9542	20	AAX20260	Borrelia burgdorfe
c	4	32.4	36.0	4015	22	AAS01490	Human secreted pro
c	5	31.6	35.1	6250	21	AAA62306	Caenorhabditis ele
c	6	31.2	34.7	3337	17	AAT34620	P. vivax ESP-1 blo
c	7	31.2	34.7	3337	20	AAX15174	DNA encoding a sec
c	8	31.2	34.7	3337	22	AAH76457	Plasmodium vivax E
	9	30.8	34.2	1226	21	AAC35351	Arabidopsis thalia
c	10	30.8	34.2	13830	22	AAD02659	Tomato chromosome
c	11	30.6	34.0	936	22	AAF58252	Oligonucleotide D1
c	12	30.6	34.0	936	22	AAF58254	Oligonucleotide D1
c	13	30.6	34.0	936	22	AAF58257	Oligonucleotide D1
c	14	30.6	34.0	936	22	AAF58259	Oligonucleotide D2
c	15	30.6	34.0	936	22	AAF58262	Oligonucleotide D2
c	16	30.6	34.0	938	22	AAF58255	Oligonucleotide D1
c	17	30.6	34.0	26928	20	AAZ32184	Human prothrombin
	18	30.6	34.0	87350	18	AAX83003	Human WRN genomic
	19	30.4	33.8	1347	20	AAX13328	Enterococcus faeca
c	20	30.4	33.8	2364	21	AAA70246	Plasmodium falcipa
	21	30	33.3	90	20	AAX60301	Fragment of the be
	22	30	33.3	244	22	AAF58238	Oligonucleotide D1
	23	30	33.3	4140	18	AAT77331	Solanum tuberosum
	24	29.8	33.1	210	20	AAX61492	B. burgdorferi ant
	25	29.8	33.1	324	20	AAX61491	B. burgdorferi ant
	26	29.8	33.1	784	22	AAH05038	Human cDNA clone (
	27	29.8	33.1	3454	22	AAH18466	Human cDNA sequenc
	28	29.8	33.1	14417	22	AAI62923	Human genomic DNA
	29	29.8	33.1	14426	22	AAI62921	Human genomic DNA
c	30	29.8	33.1	513445	22	AAI61373	Soybean 318013 reg
c	31	29.6	32.9	1518	17	AAT37313	Aromatic acyl tran
c	32	29.6	32.9	1592	21	AAZ97085	Human secreted pro
	33	29.6	32.9	1703	12	AAQ15131	Clone pACC1 encodi
	34	29.6	32.9	1703	19	AAV15701	Zucchini ACC synth
	35	29.6	32.9	1703	22	AAD04541	Zucchini 1-aminocy
	36	29.6	32.9	1703	22	AAF23618	Zucchini ACC synth
c	37	29.6	32.9	5994	21	AAA70222	Plasmodium falcipa
	38	29.6	32.9	110000	22	AAF84800	Nucleotide sequenc
	39	29.4	32.7	393	21	AAC09962	Human secreted pro
c	40	29.4	32.7	546	22	AAH09430	Human cDNA clone (
	41	29.4	32.7	1219	21	AAC48516	Arabidopsis thalia
	42	29.4	32.7	1222	21	AAC45059	Arabidopsis thalia
	43	29.4	32.7	1223	21	AAC40150	Arabidopsis thalia
c	44	29.4	32.7	1405	22	AAH34185	Human colon cancer
c	45	29.4	32.7	1978	22	AAH14169	Human cDNA sequenc

ALIGNMENTS

RESULT 1
AAX60299
ID AAX60299 standard; DNA; 1392 BP.
XX
AC AAX60299;
XX
DT 12-AUG-1999 (first entry)
XX
DE DNA encoding the beta-2 toxin of Clostridium perfringens type C.
XX
KW Beta-2 toxin; Clostridium perfringens type C; gene promoter;
KW vaccine; Clostridium tetani; ss.
XX
OS Clostridium perfringens.
XX
PN FR2768747-A1.
XX
PD 26-MAR-1999.
XX
PF 19-SEP-1997; 97FR-0011710.
XX
PR 19-SEP-1997; 97FR-0011710.
XX
PA (INSP) INST PASTEUR.
XX
PI Gibert M, Popoff MR;
XX
DR WPI; 1999-217498/19.
DR P-PSDB; AAY16591.
XX
PT Clostridium beta2 toxin gene promoter and signal sequence - useful
PT against toxins from Clostridium perfringens
XX
PS Example A; Page 31; 46pp; French.

```
XX
CC   The present sequence encodes the beta-2 toxin of Clostridium
CC   perfringens type C. The specification describes the Clostridium
CC   perfringens beta 2 toxin gene promoter (see AAX60300). The
CC   sequences can be used to produce vaccines against Clostridium,
CC   and especially Clostridium perfringens, or Clostridium
CC   tetani.
XX
SQ   Sequence 1392 BP; 606 A; 115 C; 209 G; 462 T; 0 other;


  Query Match             100.0%;  Score 90;  DB 20;  Length 1392;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-13;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 327

Qy     61 ggagcaataagtccaatgaaagcaagtgca 90
          ||||||||||||||||||||||||||||||
Db    328 ggagcaataagtccaatgaaagcaagtgca 357


RESULT   2
AAX60300
ID   AAX60300 standard; DNA; 327 BP.
XX
AC   AAX60300;
XX
DT   12-AUG-1999  (first entry)
XX
DE   Promoter of the beta-2 toxin gene of Clostridium perfringens type C.
XX
KW   Beta-2 toxin; Clostridium perfringens type C; gene promoter;
KW   vaccine; Clostridium tetani; ss.
XX
OS   Clostridium perfringens.
XX
PN   FR2768747-A1.
XX
PD   26-MAR-1999.
XX
PF   19-SEP-1997;   97FR-0011710.
XX
PR   19-SEP-1997;   97FR-0011710.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gibert M,  Popoff MR;
XX
DR   WPI; 1999-217498/19.
XX
PT   Clostridium beta2 toxin gene promoter and signal sequence - useful
PT   against toxins from Clostridium perfringens
XX
PS   Claim 1; Page 32; 46pp; French.
XX
CC   The present sequence represents the promoter of the beta-2 toxin
CC   gene of Clostridium perfringens type C. The beta2-toxin promoter
CC   and gene sequences can be used to produce vaccines against Clostridium,
CC   and especially Clostridium perfringens, or Clostridium
CC   tetani.
XX
SQ   Sequence 327 BP; 141 A; 13 C; 44 G; 129 T; 0 other;


  Query Match             66.7%;  Score 60;  DB 20;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-06;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 327


RESULT   3
AAX20260/c
ID   AAX20260 standard; DNA; 9542 BP.
XX
AC   AAX20260;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #13.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12764.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Clayton R,  Dougherty BA,  Fraser C,  Lathigra R,  Smith HO;
PI   White OR;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
PS   Claim 1; Page 920-925; 1128pp; English.
XX
CC   AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 9542 BP; 3812 A; 1160 C; 1113 G; 3457 T; 0 other;


  Query Match             36.2%;  Score 32.6;  DB 20;  Length 9542;
  Best Local Similarity   60.9%;  Pred. No. 22;
  Matches   53;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgga 63
          |||| ||  ||  | || || | | | || || || |||||||  ||| | ||   | |
Db   3932 AAAATAAACATAGCTAAATTAATTTTTATATTAATATTTTCATTATTTGTAATATCTTGT 3873

Qy     64 gcaataagtccaatgaaagcaagtgca 90
          | | ||  |  ||| |||   || |||
Db   3872 GAACTATTTATAATAAAAAGGAGAGCA 3846


RESULT   4
AAS01490/c
ID   AAS01490 standard; DNA; 4015 BP.
```

XX
AC AAS01490;
XX
DT 18-JUL-2001 (first entry)
XX
DE *Human* secreted protein gene #31.
XX
KW Human secreted protein; gene therapy; autoimmune disease;
KW hyperproliferative disorder; cardiovascular disorder;
KW cerebrovascular disorder; nervous system disorder; infection;
KW ocular disorder; wound healing; epithelial cell proliferation;
KW skin aging; transplantation; tissue regeneration; chemotaxis;
KW food additive; preservative; ds.
XX
OS Homo sapiens.
XX
PN WO200123402-A1.
XX
PD 05-APR-2001.
XX
PF 26-SEP-2000; 2000WO-US26376.
XX
PR 27-SEP-1999; 99US-0155808.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR WPI; 2001-266138/27.
DR P-PSDB; AAU01070.
XX
PT Nucleic acids encoding 43 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 4; Page 452-453; 516pp; English.
XX
CC AAS01460-AAS01502 encode for novel human secreted proteins. The
CC invention relates to 43 novel human secreted proteins (AAU01040-AAU01082)
CC and their gene sequences which can be used in gene therapy. The secreted
CC proteins are useful to prevent, treat or ameliorate a medical condition
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC sheep. The secreted proteins are also useful in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Antibodies to
CC the secreted proteins can also be used in alleviating symptoms associated
CC with disorders and in diagnostic immunoassays e.g. radioimmunoassays or
CC enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed
CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Parkinson's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities.
XX
SQ Sequence 4015 BP; 1122 A; 712 C; 743 G; 1438 T; 0 other;

```
  Query Match             36.0%;  Score 32.4;  DB 22;  Length 4015;
  Best Local Similarity   68.2%;  Pred. No. 24;
  Matches   45;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       2 tgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttg 61
           ||| ||||  |  || |||  |||| || |||| |||  |||||| | |  |  | ||||
Db    3088 TGAGAAAAGGTTATTGAAAAGTTACAGTCATTTGTATTCTTTCATTTCTCATGTTAGTTG 3029

Qy      62 gagcaa 67
           |||| |
Db    3028 GAGCTA 3023
```

RESULT 5
AAA62306/c
ID AAA62306 standard; DNA; 6250 BP.
XX
AC AAA62306;
XX
DT 12-JAN-2001 (first entry)
XX
DE Caenorhabditis elegans pdk-1 gene.
XX
KW Caenorhabditis elegans; pdk-1b; AKT kinase; daf-18;
KW insulin signalling pathway; daf-2; age-1; insulin receptor; PI 3-kinase;
KW PKB kinase; PTEN lipid phosphatase; antidiabetic; anorectic; obesity;
KW diabetes; ds.
XX
OS Caenorhabditis elegans.
XX
PN WO200033068-A1.
XX
PD 08-JUN-2000.
XX
PF 02-DEC-1999; 99WO-US28529.
XX
PR 03-DEC-1998; 98US-0205658.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Ruvkun G, Ogg S;
XX
DR WPI; 2000-423022/36.
DR P-PSDB; AAB06177, AAB06178.
XX
PT Diagnosing and treating obesity and impaired glucose tolerance using
PT modulators of daf-18 expression and/or activity -
XX
PS Disclosure; Fig 35; 402pp; English.
XX
CC The present sequence is the genomic sequence of pdk-1 from Caenorhabditis
CC elegans. It encodes two alternatively spliced isoforms which shows
CC homology to mammalian PDK. A number of C. elegans genes have been
CC identified as homologues of genes in the mammalian insulin signalling
CC pathway. The C. elegans age-1 gene encodes a homologue of the mammalian
CC PI 3-kinase whilst daf-2 encodes a homologue of the mammalian insulin
CC receptor. The C. elegans AKT kinase and PKB kinase act downstream of
CC daf-2 and age-1, just as their mammalian homologues act downstream of
CC insulin signalling. The C. elegans PTEN lipid phosphatase homologue,
CC DAF-18, has been found to act upstream of AKT in the pathway. This
CC discovery has enabled mammalian PTEN action to be mapped to the insulin
CC signalling pathway. Conserved DAF motifs can be used to design probes to
CC identify mammalian DAF homologues and thus to identify individuals with
CC a predisposition toward the development of glucose intolerance
CC conditions, such as obesity and diabetes.
XX
SQ Sequence 6250 BP; 1972 A; 1225 C; 1140 G; 1913 T; 0 other;

```
  Query Match             35.1%;  Score 31.6;  DB 21;  Length 6250;
  Best Local Similarity   62.8%;  Pred. No. 36;
  Matches   49;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       4 aaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgga 63
           ||| ||| | ||  ||| | | ||  || |||| | |||||||| ||| | ||    | |
Db     853 AAATAAAATTTTGAAAATTATTCTACAAGTTTTTTTTTTTCATTTTTGCCTAGAATAGTT 794

Qy      64 gcaataagtccaatgaaa 81
           | || | ||  ||  |||
Db     793 GTAAAATGTGAAAAAAAA 776
```

RESULT 6
AAT34620/c

```
ID   AAT34620 standard; DNA; 3337 BP.
XX
AC   AAT34620;
XX
DT   12-NOV-1996  (first entry)
XX
DE   P. vivax ESP-1 blood stage antigen coding sequence.
XX
KW   ESP-1; blood stage antigen; diagnosis; malaria; infection;
KW   causative agent; antibody; monoclonal; polyclonal; assay; ds.
XX
OS   Plasmodium vivax (clone PvMB3.3.1).
XX
FH   Key             Location/Qualifiers
FT   Exon            1..91
FT                   /*tag=  a
FT                   /note= "encodes initial (N-terminal) sequence of
FT                         hydrophobic amino acids"
FT   intron          92..230
FT                   /*tag=  b
FT                   /note= "contains typical malaria intervening
FT                         sequence splice sites splice sites"
FT   exon            231..3197
FT                   /*tag=  c
XX
PN   US5532133-A.
XX
PD   02-JUL-1996.
XX
PF   02-JUN-1993;   93US-0072610.
XX
PR   02-JUN-1993;   93US-0072610.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW;
XX
DR   WPI; 1996-321110/32.
DR   P-PSDB; AAR98747.
XX
PT   Antibodies to Plasmodium vivax blood stage antigens - used to
PT   diagnose malaria and to determine whether P. vivax is the species
PT   responsible for infection
XX
PS   Example 4; Column 15-20; 22pp; English.
XX
CC   The present sequence encodes a species-specific Plasmodium vivax
CC   malarial antigen, PvESP-1.  The gene appears to be missing a small
CC   portion of its 5' end. This protein is secreted into the plasma of
CC   a susceptible mammalian host after infection.  Monoclonal/polyclonal
CC   antibodies can be utilised in assays used to diagnose malaria, as well
CC   as to determine whether P. vivax is the species responsible for the
CC   infection.
XX
SQ   Sequence 3337 BP; 1304 A; 467 C; 875 G; 691 T; 0 other;


  Query Match             34.7%;  Score 31.2;  DB 17;  Length 3337;
  Best Local Similarity   63.2%;  Pred. No. 44;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      10 attatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaata 69
           | ||| |     | ||  | | |||||||   ||  | || ||||| ||||| ||||  |
Db     138 AATATATATGTATATATAGCAGTTTTTATAGATTATTTTTATCTTACTGTTGAAGCATGA 79

Qy      70 agtccaatgaaagcaa 85
           | |  ||  |||||||
Db      78 ATTAAAAGAAAAGCAA 63


RESULT   7
AAX15174/c
ID   AAX15174 standard; DNA; 3337 BP.
XX
AC   AAX15174;
XX
DT   28-APR-1999  (first entry)
XX
DE   DNA encoding a secreted blood-stage protein called PvESP-1.
XX
KW   Erythrocyte secreted protein-1; PvESP-1; malarial antigen;
KW   blood-stage protein; malaria; monoclonal antibody 1D11G10; ds.
XX
OS   Plasmodium vivax.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3197
FT                   /*tag=  a
FT                   /note= "contains 1 intron"
FT   exon            1..91
FT                   /*tag=  b
FT                   /number= 1
FT   intron          92..230
FT                   /*tag=  c
FT                   /number= 1
FT   exon            231..3194
FT                   /*tag=  d
FT                   /number= 2
XX
PN   US5874527-A.
XX
PD   23-FEB-1999.
XX
PF   30-SEP-1996;   96US-0719822.
XX
PR   02-JUN-1993;   93US-0072610.
PR   07-JUN-1995;   95US-0478417.
PR   30-SEP-1996;   96US-0719822.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW;
XX
DR   WPI; 1999-180063/15.
DR   P-PSDB; AAW97039.
XX
PT   Plasmodium vivax peptide antigen - for diagnosis of malaria caused
PT   by Plasmodium vivax
XX
PS   Example 4; Fig 5A-C; 23pp; English.
XX
CC   The present sequence encodes a C-terminal erythrocyte secreted
CC   protein-1 (PvESP-1) of Plasmodium vivax. PvESP-1 is a malarial
CC   antigen which is a secreted blood-stage protein present in detectable
CC   amounts in biological samples from individuals infected with P. vivax.
CC   The protein comprises an epitope not present in other Plasmodium species
CC   that cause malaria in humans, and is bound by monoclonal antibody
CC   1D11G10. The peptide antigen can be used in immunoassays for diagnosis
CC   of malaria caused by P. vivax and/or can be used to produce antibodies
CC   for use in such immunoassays.
XX
SQ   Sequence 3337 BP; 1304 A; 467 C; 875 G; 691 T; 0 other;


  Query Match             34.7%;  Score 31.2;  DB 20;  Length 3337;
  Best Local Similarity   63.2%;  Pred. No. 44;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      10 attatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaata 69
           | ||| |     | ||  | | |||||||   ||  | || ||||| ||||| ||||  |
Db     138 AATATATATGTATATATAGCAGTTTTTATAGATTATTTTTATCTTACTGTTGAAGCATGA 79

Qy      70 agtccaatgaaagcaa 85
           | |  ||  |||||||
Db      78 ATTAAAAGAAAAGCAA 63
```

```
RESULT   8
AAH76457/c
ID   AAH76457 standard; DNA; 3337 BP.
XX
AC   AAH76457;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Plasmodium vivax ESP-1 DNA.
XX
KW   Plasmodium vivax; ESP-1; erythrocyte secreted protein-1; PvESP-1;
KW   species-specific; malarial peptide antigen; infection; diagnosis;
KW   malaria; ds.
XX
OS   Plasmodium vivax.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3197
FT                   /*tag= a
FT                   /product= "ESP-1"
FT   exon            1..91
FT                   /*tag= b
FT                   /number= 1
FT   intron          92..230
FT                   /*tag= c
FT                   /number= 1
FT   exon            231..3197
FT                   /*tag= d
FT                   /number= 2
XX
PN   US6231861-B1.
XX
PD   15-MAY-2001.
XX
PF   05-JUN-1998;   98US-0092458.
XX
PR   02-JUN-1993;   93US-0072610.
PR   07-JUN-1995;   95US-0478417.
PR   30-SEP-1996;   96US-0719822.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW;
XX
DR   WPI; 2001-335068/35.
DR   P-PSDB; AAG66528.
XX
PT   New species-specific Plasmodium vivax malarial peptide antigens,
PT   proteins or fragments secreted into the plasma of susceptible mammalian
PT   host after infection, useful for diagnosing malaria  -
XX
PS   Example 4; Fig 5; 23pp; English.
XX
CC   The invention relates to novel species-specific Plasmodium vivax
CC   malarial peptide antigens which are proteins or fragments of
CC   proteins secreted into the plasma of a susceptible mammalian host after
CC   infection, and to monoclonal or polyclonal antibodies directed against
CC   those antigens. The peptide antigens, monoclonal antibodies, and/or
CC   polyclonal antibodies are useful in assays to diagnose malaria, and to
CC   determine which P. vivax species is responsible for the infection.
CC   The present sequence encodes P. vivax erythrocyte secreted
CC   protein-1 (PvESP-1), a secreted species-specific blood stage
CC   antigen provided in the invention.
XX
SQ   Sequence 3337 BP; 1304 A; 467 C; 875 G; 691 T; 0 other;


  Query Match             34.7%;  Score 31.2;  DB 22;  Length 3337;
  Best Local Similarity   63.2%;  Pred. No. 44;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     10 attatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaata 69
          | |||  |     | ||  | | |||||||   ||  | || ||||| ||||| ||||  |
Db    138 AATATATATGTATATATAGCAGTTTTTATAGATTATTTTTATCTTACTGTTGAAGCATGA 79

Qy     70 agtccaatgaaagcaa 85
          | |  ||  |||||||
Db     78 ATTAAAAGAAAAGCAA 63


RESULT   9
AAC35351
ID   AAC35351 standard; DNA; 1226 BP.
XX
AC   AAC35351;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 9873.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
```

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 34.2%; Score 30.8; DB 21; Length 1226;
Best Local Similarity 70.7%; Pred. No. 53;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 29 taatttttatgttttcatgttttcttattgttggagcaataagtccaatgaaagcaag 86
Db 1157 ttatttttttttttcatgttttctgtaaattgaagcagtaagtattaatacagcgag 1214

```
RESULT  10
AAD02659/c
ID   AAD02659 standard; DNA; 13830 BP.
XX
AC   AAD02659;
XX
DT   02-MAY-2001  (first entry)
XX
DE   Tomato chromosome 5 harbouring the RIN and MC genes.
XX
KW   Tomato; RIN; ripening inhibitor; MC; macrocalyx; sepal development;
KW   senescence; pathogen infection; ethylene response; transgenic plant; ds.
XX
OS   Lycopersicon esculentum.
XX
FH   Key             Location/Qualifiers
FT   misc_signal     1
FT                   /*tag= a
FT                   /note= "Putative transcription start site of RIN"
FT   misc_feature    1..5694
FT                   /*tag= b
FT                   /note= "Transcribed region of RIN gene"
FT   CDS             211..5489
FT                   /*tag= c
FT                   /product= "Tomato ripening-inhibitor (RIN) protein"
FT                   /note= "The specification states that the RIN gene
FT                   has 9 exons and 8 introns, however the sequence
FT                   represented in the figure 7 shows 10 exons and
FT                   9 introns"
FT   exon            211..395
FT                   /*tag= d
FT                   /number= "1"
FT   intron          396..3268
FT                   /*tag= e
FT                   /number= "1"
FT   exon            3269..3347
FT                   /*tag= f
FT                   /number= "2"
FT   intron          3348..3834
FT                   /*tag= g
FT                   /number= "2"
FT   exon            3835..3897
FT                   /*tag= h
FT                   /number= "3"
FT   intron          3898..3919
FT                   /*tag= i
FT                   /number= "3"
FT   exon            3920..3923
FT                   /*tag= j
FT                   /number= "4"
FT   intron          3924..4149
FT                   /*tag= k
FT                   /number= "4"
FT   exon            4150..4248
FT                   /*tag= l
FT                   /number= "5"
FT   intron          4249..4389
FT                   /*tag= m
FT                   /number= "5"
FT   exon            4390..4432
FT                   /*tag= n
FT                   /number= "6"
FT   intron          4433..4531
FT                   /*tag= o
FT                   /number= "6"
FT   exon            4532..4572
FT                   /*tag= p
FT                   /number= "7"
FT   intron          4573..4599
FT                   /*tag= q
FT                   /number= "7"
FT   exon            4600..4603
FT                   /*tag= r
FT                   /number= "8"
FT   intron          4604..4862
FT                   /*tag= s
FT                   /number= "8"
FT   exon            4863..4992
FT                   /*tag= t
FT                   /number= "9"
FT   intron          4993..5404
FT                   /*tag= u
FT                   /number= "9"
FT                   /note= "The rin mutation begins at a point within
FT                   this region"
FT   exon            5405..5489
FT                   /*tag= v
FT                   /number= "10"
FT   promoter        5695..8237
FT                   /*tag= w
FT                   /note= "MC promoter sequence; This region separates
FT                   the RIN from the MC transcribed region"
FT   misc_signal     8238
FT                   /*tag= x
FT                   /note= "Putative transcription start site of MC"
FT   misc_signal     8238..13830
FT                   /*tag= y
FT                   /note= "Transcribed region of MC gene"
FT   CDS             8251..13552
FT                   /*tag= z
FT                   /product= "Tomato macrocalyx (MC) protein"
FT                   /note= "The coding region has 8 exons and is interrupted
FT                   by 7 introns"
FT   exon            8251..8439
FT                   /*tag= aa
FT                   /number= "1"
FT   intron          8440..10594
FT                   /*tag= ab
FT                   /number= "1"
FT                   /note= "The rin mutation terminates within this
FT                   region of the MC gene"
FT   exon            10595..10671
FT                   /*tag= ac
FT                   /number= "2"
FT   intron          10672..11375
FT                   /*tag= ad
FT                   /number= "2"
FT   exon            11376..11445
FT                   /*tag= ae
FT                   /number= "3"
FT   intron          11446..11686
FT                   /*tag= af
FT                   /number= "3"
FT   exon            11687..11785
FT                   /*tag= ag
FT                   /number= "4"
FT   intron          11786..12163
FT                   /*tag= ah
FT                   /number= "4"
FT   exon            12164..12206
FT                   /*tag= ai
FT                   /number= "5"
FT   intron          12206..12559
FT                   /*tag= aj
FT                   /number= "5"
FT   exon            12560..12600
FT                   /*tag= ak
FT                   /number= "6"
FT   intron          12601..12879
FT                   /*tag= al
FT                   /number= "6"
FT   exon            12880..12992
FT                   /*tag= am
FT                   /number= "7"
FT   intron          12993..13449
```

```
FT                   /*tag=  an
FT                   /number= "7"
FT   exon            13450..13552
FT                   /*tag=  ao
FT                   /number= "8"
XX
PN   WO200104315-A2.
XX
PD   18-JAN-2001.
XX
PF   12-JUL-2000; 2000WO-US19035.
XX
PR   12-JUL-1999;   99US-0143364.
XX
PA   (TEXA ) UNIV TEXAS A & M SYSTEM.
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Giovannoni J,  Tanksley S,  Vrebalov J,  Padmanabhan V,  Ruezinsky D;
PI   White R;
XX
DR   WPI; 2001-103084/11.
XX
PT   New isolated nucleic acid sequence comprising RIN (ripening-inhibitor)
PT   or MC (macrocalyx) genes for use in genetic transformation techniques
PT   to manipulate a variety of plant characteristics  -
XX
PS   Disclosure; Fig 7; 167pp; English.
XX
CC   The present sequence is tomato chromosome 5 harbouring the RIN and MC
CC   genes. The invention relates to the RIN (ripening-inhibitor) and MC
CC   (macrocalyx) genes. The RIN and MC genes are useful in controlling
CC   of fruit ripening and quality, control of sepal development and
CC   size, control of senescence, control of pathogen infection, control
CC   of ethylene response, and DNA markers for assisted breeding. The
CC   RIN and MC genes are used in genetic transformation techniques to
CC   manipulate a variety of plant characteristics. Hence these genes
CC   represent a valuable new tool for the creation of transgenic plants,
CC   preferably having one or more added beneficial characteristics.
XX
SQ   Sequence 13830 BP; 4991 A; 1594 C; 1910 G; 5335 T; 0 other;


  Query Match             34.2%;  Score 30.8;  DB 22;  Length 13830;
  Best Local Similarity   58.9%;  Pred. No. 57;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          ||  || ||||||  | |||  | || |    ||  |||  |||||  ||||   ||| |
Db   7188 ATATAAGAAATTACTACAATTTATAATACTTTTCCTATAATTTCAAATTTTTAAATTTTC 7129

Qy     61 ggagcaataagtccaatgaaagcaagtgca 90
           ||   || |  |||||  || | | | ||
Db   7128 TGAAATATTAAACCAATCTAATCTAATTCA 7099


RESULT  11
AAF58252/c
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;


  Query Match             34.0%;  Score 30.6;  DB 22;  Length 936;
  Best Local Similarity   3.7%;  Pred. No. 59;
  Matches    3;  Conservative   62;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          :: :::::::::::::: ::: :::: : :::::::::: :::: :: :::: ::::: ::
Db    456 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 397

Qy     61 ggagcaataagtccaatgaaa 81
            |||::::: :  ::: :::
Db    396 TAAGCWWWWWWWWWWWWWWW 376


RESULT  12
AAF58254/c
ID   AAF58254 standard; DNA; 936 BP.
XX
AC   AAF58254;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1875.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
```

```
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;


  Query Match             34.0%;  Score 30.6;  DB 22;  Length 936;
  Best Local Similarity   3.7%;  Pred. No. 59;
  Matches    3;  Conservative   62;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           :: :::::::::::::: ::: :::: : :::::::::: :::: :: :::: ::::: ::
Db     456 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 397

Qy      61 ggagcaataagtccaatgaaa 81
             |||::::: :  ::: :::
Db     396 TAAGCWWWWWWWWWWWWWWWW 376


RESULT  13
AAF58257/c
ID   AAF58257 standard; DNA; 936 BP.
XX
AC   AAF58257;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1954.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;


  Query Match             34.0%;  Score 30.6;  DB 22;  Length 936;
  Best Local Similarity   3.7%;  Pred. No. 59;
  Matches    3;  Conservative   62;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           :: :::::::::::::: ::: :::: : :::::::::: :::: :: :::: ::::: ::
Db     456 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 397

Qy      61 ggagcaataagtccaatgaaa 81
             |||::::: :  ::: :::
Db     396 TAAGCWWWWWWWWWWWWWWWW 376


RESULT  14
AAF58259/c
ID   AAF58259 standard; DNA; 936 BP.
XX
AC   AAF58259;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2004.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;


  Query Match             34.0%;  Score 30.6;  DB 22;  Length 936;
  Best Local Similarity   3.7%;  Pred. No. 59;
  Matches    3;  Conservative   62;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           :: :::::::::::::: ::: :::: : :::::::::: :::: :: :::: ::::: ::
Db     456 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 397

Qy      61 ggagcaataagtccaatgaaa 81
             |||::::: :  ::: :::
Db     396 TAAGCWWWWWWWWWWWWWWWW 376


RESULT  15
```

AAF58262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 34.0%; Score 30.6; DB 22; Length 936;
Best Local Similarity 3.7%; Pred. No. 59;
Matches 3; Conservative 62; Mismatches 16; Indels 0; Gaps 0;

Qy 1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
:: :::::::::::::: ::: :::: : :::::::::: :::: :: :::: ::::: ::
Db 456 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 397

Qy 61 ggagcaataagtccaatgaaa 81
|||::::: : ::: :::
Db 396 TAAGCWWWWWWWWWWWWWWW 376

Search completed: January 24, 2002, 02:22:23
Job time: 1891 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:24:45 ; Search time 1502.85 Seconds
(without alignments)
987.954 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		90	100.0	90	6	AX004615	AX004615 Sequence
2		90	100.0	1392	1	L77965	L77965 Clostridium
3		90	100.0	1392	6	AX004613	AX004613 Sequence
4		88.4	98.2	54310	1	AP003515	AP003515 Clostridi
5		60	66.7	327	6	AX004614	AX004614 Sequence
6	c	37.8	42.0	128809	9	AC021648	AC021648 Homo sapi
7		36.4	40.4	152816	2	AC079739	AC079739 Homo sapi
8	c	36.4	40.4	165176	2	AC022199	AC022199 Homo sapi
9	c	36.4	40.4	167322	2	AC016788	AC016788 Homo sapi
10	c	36.4	40.4	189771	2	AP003462	AP003462 Homo sapi
11	c	36.4	40.4	193415	2	AC074180	AC074180 Homo sapi
12		35.4	39.3	148	8	CNS01BGA	AL114194 Botrytis
13	c	35.4	39.3	33460	3	AF099924	AF099924 Caenorhab
14	c	35.4	39.3	69431	3	AF067612	AF067612 Caenorhab
15	c	35.4	39.3	132319	9	AC006364	AC006364 Homo sapi
16	c	35.4	39.3	172766	2	AC079401	AC079401 Homo sapi
17		35.4	39.3	276829	2	AC006741	AC006741 Caenorhab
18	c	35.4	39.3	299864	2	AC006702	AC006702 Caenorhab
19		35	38.9	139927	9	AL354985	AL354985 Human DNA
20		34.8	38.7	154478	8	AP000423	AP000423 Arabidops
21	c	34.6	38.4	16005	3	AF049132	AF049132 Florometr
22	c	34.6	38.4	29989	3	U40059	U40059 Caenorhabdi
23		34.6	38.4	142805	2	AC022655	AC022655 Homo sapi
24	c	34.6	38.4	179633	2	AC017017	AC017017 Homo sapi
25	c	34.4	38.2	138781	9	AL157821	AL157821 Human DNA
26		34.4	38.2	155410	2	AC016459	AC016459 Homo sapi
27	c	34.2	38.0	185591	2	AC016878	AC016878 Homo sapi
28		34.2	38.0	185702	2	AP002963	AP002963 Homo sapi
29		34.2	38.0	188517	2	AP003973	AP003973 Homo sapi
30	c	34.2	38.0	200875	2	AL591711	AL591711 Mus muscu
31	c	34.2	38.0	225016	2	AC063967	AC063967 Mus muscu
32	c	34.2	38.0	269661	2	AC023911	AC023911 Homo sapi
33	c	34	37.8	176343	9	CNS01DX3	AL139021 Human chr
34		33.8	37.6	9091	3	AF324831	AF324831 Plasmodiu
35		33.8	37.6	204759	2	AC013285	AC013285 Homo sapi
36	c	33.8	37.6	245802	2	AC006279	AC006279 Plasmodiu
37		33.6	37.3	193520	2	AL589787	AL589787 Homo sapi
38	c	33.6	37.3	204652	2	PFMAL13P6	AL049183 Plasmodiu
39		33.4	37.1	2108	3	PFG27	X84904 Plasmodium
40		33.4	37.1	8597	3	PFA271088	AJ271088 Plasmodiu
41		33.4	37.1	88695	2	AC017727	AC017727 Drosophil
42		33.4	37.1	120863	3	AC092399	AC092399 Drosophil
43		33.4	37.1	163359	2	AC025699	AC025699 Homo sapi
44		33.4	37.1	175765	2	AC021111	AC021111 Homo sapi
45	c	33.4	37.1	193873	2	AC009775	AC009775 Homo sapi

ALIGNMENTS

RESULT 1
AX004615
LOCUS AX004615 90 bp DNA PAT 24-AUG-2000
DEFINITION Sequence 3 from Patent WO9915669.
ACCESSION AX004615
VERSION AX004615.1 GI:9928056
KEYWORDS .
SOURCE Clostridium perfringens.
ORGANISM Clostridium perfringens
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE 1 (bases 1 to 90)
AUTHORS Gibert,M. and Popoff,M.R.
TITLE Clostridium toxin, and method for preparing immunogenic compositions
JOURNAL Patent: WO 9915669-A 3 01-APR-1999;
GIBERT MARYSE (FR); PASTEUR INSTITUT (FR)
FEATURES Location/Qualifiers
source 1. .90
/organism="Clostridium perfringens"

```
                     /db_xref="taxon:1502"
     CDS             1. .>90
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC04902.1"
                     /db_xref="GI:9928057"
                     /translation="MKKIISKFTVIFMFSCFLIVGAISPMKASA"
BASE COUNT       31 a      9 c     14 g     36 t
ORIGIN


  Query Match             100.0%;  Score 90;  DB 6;  Length 90;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-10;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAAAAAAATTATTTCAAAGTTTACTGTAATTTTTATGTTTTCATGTTTTCTTATTGTT 60

Qy     61 ggagcaataagtccaatgaaagcaagtgca 90
          ||||||||||||||||||||||||||||||
Db     61 GGAGCAATAAGTCCAATGAAAGCAAGTGCA 90


RESULT   2
L77965
LOCUS       L77965       1392 bp    DNA             BCT       28-JUL-1998
DEFINITION  Clostridium perfringens C beta 2 toxin gene, complete cds.
ACCESSION   L77965
VERSION     L77965.1  GI:3342214
KEYWORDS    .
SOURCE      Clostridium perfringens C.
  ORGANISM  Clostridium perfringens C
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 1392)
  AUTHORS   Gibert,M., Jolivet-Reynaud,C. and Popoff,M.R.
  TITLE     Beta2 toxin, a novel toxin produced by Clostridium perfringens
  JOURNAL   Gene 203 (1), 65-73 (1997)
  MEDLINE   98085977
  REMARK    Erratum:[[published erratum appears in Gene 1998 Mar
            27;210(1):173]]
REFERENCE   2  (bases 1 to 1392)
  AUTHORS   Popoff,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1998) Toxines Microbiennes, Institut Pasteur,
            Paris cedex 15 75724, France
COMMENT     GSDB:S:76036.
            [Flatfile retrieved from GSDB Fri Jul 24 15:39:17 1998].
FEATURES             Location/Qualifiers
     source          1. .1392
                     /organism="Clostridium perfringens C"
                     /strain="CWC245"
                     /db_xref="taxon:79668"
     RBS             255. .260
                     /note="putative"
     sig_peptide     268. .357
                     /note="putative"
     CDS             268. .1065
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="beta 2 toxin"
                     /protein_id="AAC27654.1"
                     /db_xref="GI:3342215"
                     /translation="MKKIISKFTVIFMFSCFLIVGAISPMKASAKEIDAYRKVMENYL
                     NALKNYDINTVVNISEDERVNNVEQYREMLEDFKYDPNQQLKSFEILNSQKSDNKEIF
                     NVKTEFLNGAIYDMEFTVSSKDGKLIVSDMERTKVENEGKYILTPSFRTQVCTWDDEL
                     AQAIGGVYPQTYSDRFTYYADNILLNFRQYATSGSRDLKVEYSVVDHWMWKDDVKASQ
                     MVYGQNPDSARQIRLYIEKGQSFYKYRIRIKNFTPASIRVFGEGYCA"
     mat_peptide     358. .1062
                     /note="putative"
     terminator      1066. .1104
                     /note="putative"
BASE COUNT       606 a    115 c    209 g    462 t
ORIGIN


  Query Match             100.0%;  Score 90;  DB 1;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-10
  Matches   90;  Conservative    0;  Mismatches    0

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268 ATGAAAAAAATTATTTCAAAGTTTACTGTAATTTTTATGTTTTCATGTTTTCTTATTGTT 327

Qy     61 ggagcaataagtccaatgaaagcaagtgca 90
          ||||||||||||||||||||||||||||||
Db    328 GGAGCAATAAGTCCAATGAAAGCAAGTGCA 357


RESULT   3
AX004613
LOCUS       AX004613     1392 bp    DNA             PAT       24-AUG-2000
DEFINITION  Sequence 1 from Patent WO9915669.
ACCESSION   AX004613
VERSION     AX004613.1  GI:9928053
KEYWORDS    .
SOURCE      Clostridium perfringens.
  ORGANISM  Clostridium perfringens
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 1392)
  AUTHORS   Gibert,M. and Popoff,M.R.
  TITLE     Clostridium toxin, and method for preparing immunogenic
            compositions
  JOURNAL   Patent: WO 9915669-A 1 01-APR-1999;
            GIBERT MARYSE (FR); PASTEUR INSTITUT (FR)
FEATURES             Location/Qualifiers
     source          1. .1392
                     /organism="Clostridium perfringens"
                     /db_xref="taxon:1502"
     CDS             268. .1065
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC04901.1"
                     /db_xref="GI:9928054"
                     /translation="MKKIISKFTVIFMFSCFLIVGAISPMKASAKEIDAYRKVMENYL
                     NALKNYDINTVVNISEDERVNNVEQYREMLEDFKYDPNQQLKSFEILNSQKSDNKEIF
                     NVKTEFLNGAIYDMEFTVSSKDGKLIVSDMERTKVENEGKYILTPSFRTQVCTWDDEL
                     AQAIGGVYPQTYSDRFTYYADNILLNFRQYATSGSRDLKVEYSVVDHWMWKDDVKASQ
                     MVYGQNPDSARQIRLYIEKGQSFYKYRIRIKNFTPASIRVFGEGYCA"
BASE COUNT       606 a    115 c    209 g    462 t
ORIGIN


  Query Match             100.0%;  Score 90;  DB 6;  Length 1392;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-10;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268 ATGAAAAAAATTATTTCAAAGTTTACTGTAATTTTTATGTTTTCATGTTTTCTTATTGTT 327

Qy     61 ggagcaataagtccaatgaaagcaagtgca 90
          ||||||||||||||||||||||||||||||
Db    328 GGAGCAATAAGTCCAATGAAAGCAAGTGCA 357


RESULT   4
AP003515
LOCUS       AP003515    54310 bp    DNA   circular  BCT       10-AUG-2001
DEFINITION  Clostridium perfringens plasmid pCP13 DNA, complete sequence.
```

```
ACCESSION   AP003515
VERSION     AP003515.1  GI:15076712
KEYWORDS    .
SOURCE      Clostridium perfringens (strain:13) plasmid:pCP13 DNA.
  ORGANISM  Clostridium perfringens
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 54310)
  AUTHORS   Ohtani,K., Ohshima,S., Hirakawa,H., Ohshima,K., Shiba,T.,
            Shimizu,T., Hattori,M., Kuhara,S., Hayashi,H. and Shimizu,T.
  TITLE     Complete Nucleotide Sequence of the Virulence Plasmid pCP13 from
            Clostridium perfringens
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 54310)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2001) Tohru Shimizu, Institute of Basic Medical
            Sciences, University of Tukuba, Department of Microbiology; 1-1-1
            Tennohdai, Tsukuba, Ibaraki 305-8575, Japan
            (E-mail:tshimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
            Fax:81-298-53-3354)
FEATURES             Location/Qualifiers
     source          1. .54310
                     /organism="Clostridium perfringens"
                     /plasmid="pCP13"
                     /strain="13"
                     /db_xref="taxon:1502"
                     /note="anaerobic pathogen for gas gangrene"
     gene            940. .1692
                     /gene="soj"
     CDS             940. .1692
                     /gene="soj"
                     /note="250 aa, similar to pir:I40444 Spo0A activation
                     inhibitor soj from Bacillus subtilis (253 aa); 37%
                     identity in 250 aa overlap
                     PCP01
                     ParA family"
                     /codon_start=1
                     /transl_table=11
                     /product="Soj protein"
                     /protein_id="BAB62438.1"
                     /db_xref="GI:15076713"
                     /translation="MKKISVFNIKGGVAKTTSTANFGACLEEKGDRVLLVDLDPQSNL
                     TKLFKAYSMEDVSIADVLLDKNLDIEKVIKKTDFENIDILPSNVTLAFAERKILLDVN
                     RSQQNRLAKALEEIEDKYDYCLIDCPPALNMITVNALCASDEVLVPIKIDKFALDGLE
                     YLLDSIEEIKDEFNPNLNFKGCFITMDSSTTVNKVIKQELKSVLGEKMFNTSIHQNIK
                     VVESTFEECPVVFSSKKARASLNYKDLSKEIF"
     gene            1751. .3031
                     /gene="parB"
     CDS             1751. .3031
                     /gene="parB"
                     /note="426 aa, similar to gpu:AF300944_3 presumptive ParB
                     protein from Lactococcus lactis subsplactis (242 aa); 30%
                     identity in 266 aa overlap
                     PCP02"
                     /codon_start=1
                     /transl_table=11
                     /product="ParB protein"
                     /protein_id="BAB62439.1"
                     /db_xref="GI:15076714"
                     /translation="MAKKFSISEGMLNGISKNTRKVEEFQAKENFKIEYINIDRIKRN
                     EKNFYEIVDIESLAEDIKLNGLNHNLVVRKLDNDMYELISGERRYTALSKLVNEGNKE
                     FNLVPCKVIESNDIDSEIILIQANAQSRELTEVEKLTQVERLQEFYKIKKENGEKVPG
                     KIRDIIANDLNLSATQVGRYERINNKLIPELKAVIEQGNLTIANASEFSSLSEENQRV
                     ILSIIDDKTNMSKQEAVDLKNKLKKIEEEKELELKKAYEEKELELKRLEEEKKNQVSK
                     LKSENENLKKKLDSNNIEEERKEIEGQIKIEFEEKLKNEKVILEEELKSKYDKKIEDI
                     TKEAKENNLEKQRLKDELSKLKEKSNNEVDIKNTKENFVLVQNLKLIDNSFKNLKSQI
                     NKMKKENVKVAEETKAKEFLEKYQKEISDLLKKL"
     gene            3147. .3509
                     /gene="PCP03"
     CDS             3147. .3509
                     /gene="PCP03"
                     /note="120 aa, no significant homology"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB62440.1"
                     /db_xref="GI:15076715"
                     /translation="MEKILAEKRINISFYKRKNGALVTTLYLPPKWLEVIGITENERE
                     CFFYIEDKAIKISKEKQSEEAKEKTISFSKTSTKTYLNNKWLEYLGVSEDERSCIIEL
                     RKKDITLVKDNGRDILDI"
     gene            3773. .4024
                     /gene="PCP04"
     CDS             3773. .4024
                     /gene="PCP04"
                     /note="83 aa, similar to pir:T14710 probable transposase
                     from Yersinia pestis (402 aa); 44% identity in 50 aa
                     overlap
                     truncated"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transposase"
                     /protein_id="BAB62442.1"
                     /db_xref="GI:15076717"
                     /translation="MSKDFLRNYIKEQNFNDSNNVLNSLKDLFNDVLQEVVSAELDDM
                     LGYEKNLTNQFYNSRNGYSKKTIKSELGTITLNILRYRN"
     gene            4040. .4222
                     /gene="PCP05"
     CDS             4040. .4222
                     /gene="PCP05"
                     /note="60 aa, similar to gp:AF143819_1 transposase-like
                     protein from Escherichia coli (402 aa); 38% identity in 60
                     aa overlap
                     truncated"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transposase"
                     /protein_id="BAB62443.1"
                     /db_xref="GI:15076718"
                     /translation="MISKHQRNINRIEDKNLNLYASGITTRDVAGQIKALYDIEISAE
                     TVSNITNRIMPLVSEW"
     gene            4588. .4746
                     /gene="PCP06"
     CDS             4588. .4746
                     /gene="PCP06"
                     /note="52 aa, similar to pir:T43600 probable transposase
                     from Yersinia pestis (105 aa); 56% identity in 50 aa
                     overlap
                     truncated"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transposase"
                     /protein_id="BAB62444.1"
                     /db_xref="GI:15076719"
                     /translation="MLYTTNVIESLNSQFRKFTKTKLIFPNDVSLIKMLYLATEKVNK
                     KWTRNYPN"
     gene            complement(5169. .5804)
                     /gene="PCP07"
     CDS             complement(5169. .5804)
                     /gene="PCP07"
                     /function="ATP-binding protein"
                     /note="211 aa, similar to gp:AP001508_4 ABC transporter
                     (ATP-binding protein) from Bacillus halodurans (213 aa);
                     49% identity in 214 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="probable ABC transporter"
                     /protein_id="BAB62445.1"
                     /db_xref="GI:15076720"
                     /translation="MNIIEISNLNKKYFDKVIFKDFSLSIKKGEMIAISGRSGCGKST
                     LLNMIGLIEKFDSGEIIIDGVKNIKINSKLANKFLREKISYLFQNFALVDEETVEENL
                     RLAIKHTIKNTKKIEEEIIRCLKFVGLEGCQKNYIYELSGGEQQRVAIARLMLKPSEI
                     ILADEPTGSLDEENRDIIISLLKELNESGKTIIIVTHDNYVAKQADRIIFL"
     gene            complement(5804. .7966)
                     /gene="PCP08"
     CDS             complement(5804. .7966)
```

```
                /gene="PCP08"
                /note="720 aa, similar to gp:AP001508_3 BH0280 gene
                product from Bacillus halodurans (713 aa); 23% identity in
                661 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="BAB62446.1"
                /db_xref="GI:15076721"
                /translation="MKKKVALIILFIIITILSFEGVYSVRNHTEFMKLKNLQGNLENF
                EVTISIPDRENYNTAYESIIKSLDEYNGNIFFSEVDINEDVRKYINYGYFSNKETENH
                IPIVSGRFFHEKDNIDSYLSTIDSEDTQQIGVINDFNGKNIHEIRTIKSKLDINTFEN
                LFIVQIENEQILDKLIEDLKSESIIVQKRVMGDSSQYNTETLKIILVVCFIGLIFMIF
                YQVLGSYKKIGIQKLLGHSTFVMLKERLLEVLRIEVIVMLVVTVLLVFFNFKTFNSLF
                WKFMLELICIYSIMIIFTIVSVIIPYIYVSKITLSNIIKNKRPVKSIIILNSIVKVIL
                ASIILIFFSNALDDLSSIGKGYEKNYKVWEETKQYYILPELGFNDESIQSFSIEEMEK
                ERAVYLYFNKQGAILADFNRYEPTSMEEAKQMLPEEYMRETIIVNPNYLLKHKVYDVD
                GNIINISEDEKDRILLVPEKYRNFEKEILEYYGYNSQEPCSSTTCSHKTADGKLNLVE
                QKQKIIWMKSNQKYFSYLLDVNPEEGNYVTDPIVSVLTESNDKLVSYYKIIGYNNSPF
                KIRANSEEEVINGLEKYYDMSVYLIDPYNLYDNVASTIINIKAKVKVIIFAIVILLAV
                ISIIILQNTSLYFNQNKNKIIVKKLHGYRLIYRYMNYFIMVLITWTCPLAIASLITKD
                INIIYFTLILVVIELVFIIFNINSLEKKNLIKVIKGEY"
     gene       complement(8043. .8354)
                /gene="PCP09"
     CDS        complement(8043. .8354)
                /gene="PCP09"
                /note="103 aa, no significant homology"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB62447.1"
                /db_xref="GI:15076722"
                /translation="MKRNKLIKTFIASLLLVGSLGTTALAYTHSSDKFEAASLPGLFS
                QAQSSKFYCGEQKHRATARVKVGTTLYEAKDIKDAKLTAHAQTKYYKGVTEWNSYYAH
                L"
     gene       8779. .9012
                /gene="PCP10"
     CDS        8779. .9012
                /gene="PCP10"
                /note="77 aa, similar to probable transposase from
                Yersinia pestis plasmid pMT1 (402 aa); 25% identity in 158
                aa overlap
                truncated"
                /codon_start=1
                /transl_table=11
                /product="probable transposase"
                /protein_id="BAB62448.1"
                /db_xref="GI:15076723"
                /translation="MKEVLSICMSTNEGSKFWLSPHKDKKEFAKDLKTIYGSVNETEG
                MKNLIELREKWGSKYPNVVKSWKDNWDNLSTFF"
     gene       9127. .9366
                /gene="PCP11"

  Query Match           98.2%;  Score 88.4;  DB 1;  Length 54310;
  Best Local Similarity  98.9%;  Pred. No. 3.9e-10;
  Matches   89;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   13654 ATGAAAAAAATTATTTCAAAGTTTACTGTAATTTTTATGTTTTCATATTTTCTTATTGTT 13713

Qy      61 ggagcaataagtccaatgaaagcaagtgca 90
           ||||||||||||||||||||||||||||||
Db   13714 GGAGCAATAAGTCCAATGAAAGCAAGTGCA 13743


RESULT   5
AX004614
LOCUS       AX004614      327 bp    DNA             PAT       24-AUG-2000
DEFINITION  Sequence 2 from Patent WO9915669.
ACCESSION   AX004614
VERSION     AX004614.1  GI:9928055
KEYWORDS    .
SOURCE      Clostridium perfringens.
  ORGANISM  Clostridium perfringens
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Gibert,M. and Popoff,M.R.
  TITLE     Clostridium toxin, and method for preparing immunogenic
            compositions
  JOURNAL   Patent: WO 9915669-A 2 01-APR-1999;
            GIBERT MARYSE (FR); PASTEUR INSTITUT (FR)
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Clostridium perfringens"
                     /db_xref="taxon:1502"
BASE COUNT      141 a     13 c     44 g    129 t
ORIGIN


  Query Match           66.7%;  Score 60;  DB 6;  Length 327;
  Best Local Similarity  100.0%;  Pred. No. 0.00053;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     268 ATGAAAAAAATTATTTCAAAGTTTACTGTAATTTTTATGTTTTCATGTTTTCTTATTGTT 327


RESULT   6
AC021648/c
LOCUS       AC021648   128809 bp    DNA             PRI       01-MAY-2001
DEFINITION  Homo sapiens 12 BAC RP11-255I14 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC021648
VERSION     AC021648.25  GI:13899174
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128809)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
```

Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 128809)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 128809)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 1, 2001 this sequence version replaced gi:13876397.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

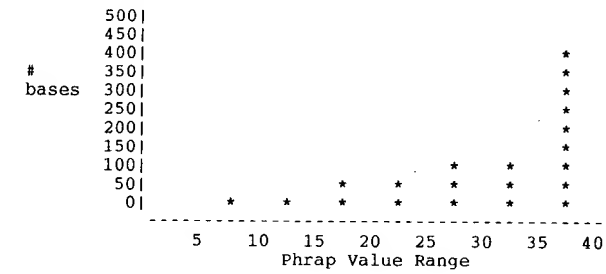
--- Summary Statistics ---

Contig length:	128809
Phrap values in estimate:	128476
Average error rate (BCM-Phrap estimate):	3.45519e-05
Fraction of Phrap values less than 40 :	0.00621906
Number of consensus changing edits:	3
Number of N's in consensus :	0

--- Consensus changing edits ---

Position	Original+Context	Edited+Context
93661	gggatattaa(n)atgggtttgt	gggatattaa(a)atgggtttgt
94679	gtctcccctta(n)ataaaatgat	gtctcccctta(a)ataaaatgat
119599	aaaaaaaaaa(n)gaattaccca	aaaaaaaaaa(a)gaattaccca

--- Distribution of Quality < 40 Bases ---



Version: 1.01 qxfo.

FEATURES	Location/Qualifiers
source	1. .128809
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="12"
	/clone="RP11-255I14"
repeat_region	complement(1. .114)
	/rpt_family="L1M4"
misc_feature	complement(1. .2005)
	/note="Overlaps bases 2005. .1 of clone AC025257"
	/function="Overlaps with adjacent clone AC025257"
repeat_region	256. .327
	/rpt_family="LTR1B"
repeat_region	264. .1339
	/rpt_family="MER52A"
repeat_region	1850. .1903
	/rpt_family="(TAA)n"
repeat_region	complement(2108. .2229)
	/rpt_family="L1MEc"
repeat_region	2258. .2281
	/rpt_family="AT_rich"
repeat_region	3464. .3505
	/rpt_family="AT_rich"
repeat_region	complement(3929. .4290)
	/rpt_family="THE1B"
repeat_region	complement(4428. .4640)
	/rpt_family="MLT1J"
repeat_region	5805. .5834
	/rpt_family="AT_rich"
repeat_region	5950. .6723
	/rpt_family="L1PA2"
repeat_region	6719. .11958
	/rpt_family="L1PA2"
STS	9223. .9436
	/standard_name="G44369"
	/db_xref="dbSTS:66785"
STS	10417. .10594
	/standard_name="G02335"
	/db_xref="dbSTS:7821"
repeat_region	11994. .12330
	/rpt_family="THE1B"
repeat_region	12920. .12976
	/rpt_family="AT_rich"
repeat_region	complement(13797. .14024)
	/rpt_family="L1M4"
repeat_region	complement(14409. .14518)
	/rpt_family="MIR"

```
     repeat_region   15194. .15627
                     /rpt_family="MLT1C"
     repeat_region   15971. .15996
                     /rpt_family="AT_rich"
     repeat_region   16346. .16396
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(16397. .16704)
                     /rpt_family="AluSg"
     repeat_region   17398. .17434
                     /rpt_family="AT_rich"

  Query Match             42.0%;  Score 37.8;  DB 9;  Length 128809;
  Best Local Similarity   64.0%;  Pred. No. 29;
  Matches   57;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           ||| ||  || | || || ||| |||| ||||||| ||  || ||  ||   ||||   |
Db   65107 ATGGAATTAAATGTTGCACAGTCTACTTTAATTTTAATAATTCCAGATTCATTTATAAAT 65048

Qy      61 ggagcaataagtccaatgaaagcaagtgc 89
           |||  ||| ||   ||||| |  | || |
Db   65047 GGAAGAATGAGAAAAATGAGAAAATGTTC 65019


RESULT   7
AC079739
LOCUS       AC079739   152816 bp    DNA             HTG       29-SEP-2000
DEFINITION  Homo sapiens chromosome UNK clone CTD-2012B22, WORKING DRAFT
            SEQUENCE, 34 unordered pieces.
ACCESSION   AC079739
VERSION     AC079739.2  GI:10305265
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152816)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152816)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi:10047917.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_MS2012B22
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 95%
            Sequencing vector: plasmid; 5%
            Chemistry: Dye-primer ET; 95% of reads
            Chemistry: Dye-terminator Big Dye; 5% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 130608 bases at least Q40
            Consensus quality: 137552 bases at least Q30
            Consensus quality: 141453 bases at least Q20
            Insert size: 150000; agarose-fp
            Insert size: 149516; sum-of-contigs
            Quality coverage: 3.13 in Q20 bases; agarose-fp
            Quality coverage: 3.30 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1155: contig of 1155 bp in length
            *     1156     1255: gap of unknown length
            *     1256     2579: contig of 1324 bp in length
            *     2580     2679: gap of unknown length
            *     2680     3782: contig of 1103 bp in length
            *     3783     3882: gap of unknown length
            *     3883     5263: contig of 1381 bp in length
            *     5264     5363: gap of unknown length
            *     5364     7272: contig of 1909 bp in length
            *     7273     7372: gap of unknown length
            *     7373     8845: contig of 1473 bp in length
            *     8846     8945: gap of unknown length
            *     8946    10943: contig of 1998 bp in length
            *    10944    11043: gap of unknown length
            *    11044    12586: contig of 1543 bp in length
            *    12587    12686: gap of unknown length
            *    12687    14738: contig of 2052 bp in length
            *    14739    14838: gap of unknown length
            *    14839    16585: contig of 1747 bp in length
            *    16586    16685: gap of unknown length
            *    16686    18458: contig of 1773 bp in length
            *    18459    18558: gap of unknown length
            *    18559    20559: contig of 2001 bp in length
            *    20560    20659: gap of unknown length
            *    20660    23709: contig of 3050 bp in length
            *    23710    23809: gap of unknown length
            *    23810    25811: contig of 2002 bp in length
            *    25812    25911: gap of unknown length
            *    25912    28324: contig of 2413 bp in length
            *    28325    28424: gap of unknown length
            *    28425    33027: contig of 4603 bp in length
            *    33028    33127: gap of unknown length
            *    33128    37234: contig of 4107 bp in length
            *    37235    37334: gap of unknown length
            *    37335    41889: contig of 4555 bp in length
            *    41890    41989: gap of unknown length
            *    41990    45247: contig of 3258 bp in length
            *    45248    45347: gap of unknown length
            *    45348    50669: contig of 5322 bp in length
            *    50670    50769: gap of unknown length
            *    50770    54317: contig of 3548 bp in length
            *    54318    54417: gap of unknown length
            *    54418    58856: contig of 4439 bp in length
            *    58857    58956: gap of unknown length
            *    58957    63667: contig of 4711 bp in length
            *    63668    63767: gap of unknown length
            *    63768    68619: contig of 4852 bp in length
            *    68620    68719: gap of unknown length
            *    68720    73963: contig of 5244 bp in length
            *    73964    74063: gap of unknown length
            *    74064    78512: contig of 4449 bp in length
            *    78513    78612: gap of unknown length
            *    78613    85002: contig of 6390 bp in length
            *    85003    85102: gap of unknown length
            *    85103    90498: contig of 5396 bp in length
            *    90499    90598: gap of unknown length
            *    90599    98523: contig of 7925 bp in length
            *    98524    98623: gap of unknown length
            *    98624   106597: contig of 7974 bp in length
            *   106598   106697: gap of unknown length
            *   106698   114088: contig of 7391 bp in length
            *   114089   114188: gap of unknown length
            *   114189   123937: contig of 9749 bp in length
            *   123938   124037: gap of unknown length
            *   124038   133862: contig of 9825 bp in length
            *   133863   133962: gap of unknown length
            *   133963   152816: contig of 18854 bp in length.
FEATURES             Location/Qualifiers
     source          1. .152816
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="CTD-2012B22"
     misc_feature    1. .1155
                     /note="assembly_name:Contig12"
     misc_feature    1256. .2579
                     /note="assembly_name:Contig15"
     misc_feature    2680. .3782
                     /note="assembly_name:Contig16"
     misc_feature    3883. .5263
                     /note="assembly_name:Contig17"
     misc_feature    5364. .7272
                     /note="assembly_name:Contig18"
     misc_feature    7373. .8845
                     /note="assembly_name:Contig19"
     misc_feature    8946. .10943
                     /note="assembly_name:Contig20"
     misc_feature    11044. .12586
                     /note="assembly_name:Contig21"
     misc_feature    12687. .14738
                     /note="assembly_name:Contig22"
     misc_feature    14839. .16585
                     /note="assembly_name:Contig23"
     misc_feature    16686. .18458
                     /note="assembly_name:Contig24"
     misc_feature    18559. .20559
                     /note="assembly_name:Contig25"
     misc_feature    20660. .23709
                     /note="assembly_name:Contig26"
     misc_feature    23810. .25811
                     /note="assembly_name:Contig27"
     misc_feature    25912. .28324
                     /note="assembly_name:Contig28"
     misc_feature    28425. .33027
                     /note="assembly_name:Contig29"
     misc_feature    33128. .37234
                     /note="assembly_name:Contig30"
     misc_feature    37335. .41889
                     /note="assembly_name:Contig31"
     misc_feature    41990. .45247
                     /note="assembly_name:Contig32"
     misc_feature    45348. .50669
                     /note="assembly_name:Contig33"
     misc_feature    50770. .54317
                     /note="assembly_name:Contig34"
     misc_feature    54418. .58856
                     /note="assembly_name:Contig35"
     misc_feature    58957. .63667
                     /note="assembly_name:Contig36"
     misc_feature    63768. .68619
                     /note="assembly_name:Contig37"
     misc_feature    68720. .73963
                     /note="assembly_name:Contig38"
     misc_feature    74064. .78512
                     /note="assembly_name:Contig39"
     misc_feature    78613. .85002
                     /note="assembly_name:Contig40"
     misc_feature    85103. .90498
                     /note="assembly_name:Contig41"
     misc_feature    90599. .98523
                     /note="assembly_name:Contig42"
     misc_feature    98624. .106597
                     /note="assembly_name:Contig43"
     misc_feature    106698. .114088
                     /note="assembly_name:Contig44"
     misc_feature    114189. .123937
                     /note="assembly_name:Contig45"
     misc_feature    124038. .133862
                     /note="assembly_name:Contig46"
     misc_feature    133963. .152816
                     /note="assembly_name:Contig47"
BASE COUNT    44029 a  29445 c  29570 g  46464 t   3308 others
ORIGIN
```

```
  Query Match             40.4%;  Score 36.4;  DB 2;  Length 152816;
  Best Local Similarity   64.0%;  Pred. No. 57;
  Matches   55;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       3 gaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgg 62
           |||||| ||||||| |||  ||||||   ||| ||| |||| |||||||  |  | ||
Db   42020 GAAAAATATTATTTAAAATATTACTGCTTTTTCTATATTTTTATGTTTTTCTGGTTTTTC 42079

Qy      63 agcaataagtccaatgaaagcaagtg 88
             | | |||     |  |||  | ||
Db   42080 TACCACAAGAATCTTATAAGTCAATG 42105


RESULT   8
AC022199/c
LOCUS       AC022199   165176 bp    DNA             HTG       05-APR-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-145G19 map 17, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC022199
VERSION     AC022199.2  GI:7417783
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165176)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-145G19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165176)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 5, 2000 this sequence version replaced gi:6759152.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5174
                Center clone name: 145_G_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 157408 bases at least Q40
                Consensus quality: 161068 bases at least Q30
                Consensus quality: 162457 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 164076; sum-of-contigs
```

```
            Quality coverage: 4.3 in Q20 bases; agarose-fp
            Quality coverage: 4.4 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     3498: contig of 3498 bp in length
          *     3499 3598: gap of      100 bp
          *     3599     6363: contig of 2765 bp in length
          *     6364 6463: gap of      100 bp
          *     6464     8738: contig of 2275 bp in length
          *     8739 8838: gap of      100 bp
          *     8839    12799: contig of 3961 bp in length
          *    12800 12899: gap of      100 bp
          *    12900    19411: contig of 6512 bp in length
          *    19412 19511: gap of      100 bp
          *    19512    29228: contig of 9717 bp in length
          *    29229 29328: gap of      100 bp
          *    29329    40493: contig of 11165 bp in length
          *    40494 40593: gap of      100 bp
          *    40594    52209: contig of 11616 bp in length
          *    52210 52309: gap of      100 bp
          *    52310    71806: contig of 19497 bp in length
          *    71807 71906: gap of      100 bp
          *    71907    98472: contig of 26566 bp in length
          *    98473 98572: gap of      100 bp
          *    98573   128201: contig of 29629 bp in length
          *   128202 128301: gap of      100 bp
          *   128302   165176: contig of 36875 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165176
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-145G19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .3498
                     /note="assembly_fragment"
     misc_feature    3599. .6363
                     /note="assembly_fragment"
     misc_feature    6464. .8738
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8839. .12799
                     /note="assembly_fragment"
     misc_feature    12900. .19411
                     /note="assembly_fragment"
     misc_feature    19512. .29228
                     /note="assembly_fragment"
     misc_feature    29329. .40493
                     /note="assembly_fragment"
     misc_feature    40594. .52209
                     /note="assembly_fragment"
     misc_feature    52310. .71806
                     /note="assembly_fragment"
     misc_feature    71907. .98472
                     /note="assembly_fragment"
     misc_feature    98573. .128201
                     /note="assembly_fragment"
     misc_feature    128302. .165176
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
BASE COUNT    51832 a  30819 c  31782 g  49641 t   1102 others
ORIGIN
```

```
  Query Match          40.4%;  Score 36.4;  DB 2;  Length 165176;
  Best Local Similarity  64.0%;  Pred. No. 57;
  Matches   55;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       3 gaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgg 62
           |||||| ||||||| |||  ||||||   ||| ||| |||| |||||||  |  | ||
Db   18382 GAAAAATATTATTTAAAATATTACTGCTTTTTCTATATTTTTATGTTTTTCTGGTTTTTC 18323

Qy      63 agcaataagtccaatgaaagcaagtg 88
           | | |||     |  |||  | ||
Db   18322 TACCACAAGAATCTTATAAGTCAATG 18297


RESULT   9
AC016788/c
LOCUS       AC016788   167322 bp    DNA             HTG       20-SEP-2000
DEFINITION  Homo sapiens clone RP11-11E12, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC016788
VERSION     AC016788.6  GI:10198423
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167322)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-11E12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167322)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 20, 2000 this sequence version replaced gi:9123832.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3141
                Center clone name: 11_E_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 159178 bases at least Q40
                Consensus quality: 163354 bases at least Q30
                Consensus quality: 164761 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 165922; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; agarose-fp
                Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
```

```
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4327: contig of 4327 bp in length
            *     4328  4427: gap of      100 bp
            *     4428     6881: contig of 2454 bp in length
            *     6882  6981: gap of      100 bp
            *     6982     9727: contig of 2746 bp in length
            *     9728  9827: gap of      100 bp
            *     9828    12474: contig of 2647 bp in length
            *    12475 12574: gap of      100 bp
            *    12575    14864: contig of 2290 bp in length
            *    14865 14964: gap of      100 bp
            *    14965    18876: contig of 3912 bp in length
            *    18877 18976: gap of      100 bp
            *    18977    60052: contig of 41076 bp in length
            *    60053 60152: gap of      100 bp
            *    60153    69381: contig of 9229 bp in length
            *    69382 69481: gap of      100 bp
            *    69482    77833: contig of 8352 bp in length
            *    77834 77933: gap of      100 bp
            *    77934    86115: contig of 8182 bp in length
            *    86116 86215: gap of      100 bp
            *    86216    98453: contig of 12238 bp in length
            *    98454 98553: gap of      100 bp
            *    98554   116358: contig of 17805 bp in length
            *   116359 116458: gap of      100 bp
            *   116459   134093: contig of 17635 bp in length
            *   134094 134193: gap of      100 bp
            *   134194   162917: contig of 28724 bp in length
            *   162918 163017: gap of      100 bp
            *   163018   167322: contig of 4305 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167322
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-11E12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .4327
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4428. .6881
                     /note="assembly_fragment"
     misc_feature    6982. .9727
                     /note="assembly_fragment"
     misc_feature    9828. .12474
                     /note="assembly_fragment"
     misc_feature    12575. .14864
                     /note="assembly_fragment"
     misc_feature    14965. .18876
                     /note="assembly_fragment"
     misc_feature    18977. .60052
                     /note="assembly_fragment"
     misc_feature    60153. .69381
                     /note="assembly_fragment"
     misc_feature    69482. .77833
                     /note="assembly_fragment"
     misc_feature    77934. .86115
                     /note="assembly_fragment"
     misc_feature    86216. .98453
                     /note="assembly_fragment"
     misc_feature    98554. .116358
                     /note="assembly_fragment"
     misc_feature    116459. .134093
                     /note="assembly_fragment"
     misc_feature    134194. .162917
                     /note="assembly_fragment"
     misc_feature    163018. .167322
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT     50484 a   31742 c   33143 g   50550 t    1403 others
ORIGIN


  Query Match             40.4%;  Score 36.4;  DB 2;  Length 167322;
  Best Local Similarity   64.0%;  Pred. No. 57;
  Matches   55;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       3 gaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgg 62
           ||||||| ||||||| |||  ||||||   ||| ||| |||| |||||||  |  | ||
Db 158772 GAAAAATATTATTTAAAATATTACTGCTTTTTCTATATTTTTATGTTTTTCTGGTTTTTC 158713

Qy      63 agcaataagtccaatgaaagcaagtg 88
             | | |||     |  |||  | ||
Db 158712 TACCACAAGAATCTTATAAGTCAATG 158687


RESULT  10
AP003462/c
LOCUS       AP003462   189771 bp    DNA             HTG       30-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-480C22 map 11q, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AP003462
VERSION     AP003462.1  GI:13488923
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-480C22.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189771)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     -------------- Genome Center
               Center: RIKEN Genomic Sciences Center(GSC)
               Center code: RIKEN
               Web site: http://hgp.gsc.riken.go.jp/
               Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
               Center project name: HumDraft11
               Center clone name: RP11-480C22
            -------------- Summary Statistics
               Sequencing vector: PCR products; 100% of reads
               Chemistry: Dye-terminator ET-amersham; 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 187154 bases at least Q40
               Consensus quality: 188004 bases at least Q30
               Consensus quality: 188284 bases at least Q20
               Insert size: 188571; sum-of-contigs
               Quality coverage: 8.56x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            13 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     43166 contig of  43166 bp in length
              43267     62703 contig of  19437 bp in length
              62804     84476 contig of  21673 bp in length
              84577    106668 contig of  22092 bp in length
             106769    126319 contig of  19551 bp in length
             126420    141481 contig of  15062 bp in length
             141582    153916 contig of  12335 bp in length
```

```
            154017    163781 contig of   9765 bp in length
            163882    170760 contig of   6879 bp in length
            170861    176864 contig of   6004 bp in length
            176965    183161 contig of   6197 bp in length
            183262    185847 contig of   2586 bp in length
            185948    189771 contig of   3824 bp in length.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 13 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    43166: contig of 43166 bp in length
          *   43167 43266: gap of      100 bp
          *   43267    62703: contig of 19437 bp in length
          *   62704 62803: gap of      100 bp
          *   62804    84476: contig of 21673 bp in length
          *   84477 84576: gap of      100 bp
          *   84577   106668: contig of 22092 bp in length
          *  106669 106768: gap of      100 bp
          *  106769   126319: contig of 19551 bp in length
          *  126320 126419: gap of      100 bp
          *  126420   141481: contig of 15062 bp in length
          *  141482 141581: gap of      100 bp
          *  141582   153916: contig of 12335 bp in length
          *  153917 154016: gap of      100 bp
          *  154017   163781: contig of 9765 bp in length
          *  163782 163881: gap of      100 bp
          *  163882   170760: contig of 6879 bp in length
          *  170761 170860: gap of      100 bp
          *  170861   176864: contig of 6004 bp in length
          *  176865 176964: gap of      100 bp
          *  176965   183161: contig of 6197 bp in length
          *  183162 183261: gap of      100 bp
          *  183262   185847: contig of 2586 bp in length
          *  185848 185947: gap of      100 bp
          *  185948   189771: contig of 3824 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189771
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-480C22"
     misc_feature    1. .43166
                     /note="assembly_fragment"
     misc_feature    43267. .62703
                     /note="assembly_fragment"
     misc_feature    62804. .84476
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    84577. .106668
                     /note="assembly_fragment"
     misc_feature    106769. .126319
                     /note="assembly_fragment"
     misc_feature    126420. .141481
                     /note="assembly_fragment"
     misc_feature    141582. .153916
                     /note="assembly_fragment"
     misc_feature    154017. .163781
                     /note="assembly_fragment"
     misc_feature    163882. .170760
                     /note="assembly_fragment"
     misc_feature    170861. .176864
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    176965. .183161
                     /note="assembly_fragment"
     misc_feature    183262. .185847
                     /note="assembly_fragment"
     misc_feature    185948. .189771
                     /note="assembly_fragment"
BASE COUNT    56729 a  37254 c  36581 g  58007 t   1200 others
```

```
ORIGIN


  Query Match             40.4%;  Score 36.4;  DB 2;  Length 189771;
  Best Local Similarity   64.0%;  Pred. No. 57;
  Matches   55;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       3 gaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgg 62
           |||||| ||||||| |||  ||||||   ||| ||| |||| |||||||  |  | ||
Db 140789 GAAAAATATTATTTAAAATATTACTGCTTTTTCTATATTTTTATGTTTTTCTGGTTTTTC 140730

Qy      63 agcaataagtccaatgaaagcaagtg 88
             | | |||     |  |||  | ||
Db 140729 TACCACAAGAATCTTATAAGTCAATG 140704


RESULT  11
AC074180/c
LOCUS       AC074180   193415 bp    DNA              HTG       17-AUG-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-480C22, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC074180
VERSION     AC074180.2  GI:9838218
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193415)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193415)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 17, 2000 this sequence version replaced gi:9211440.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0480C22
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 2%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 2% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 187074 bases at least Q40
            Consensus quality: 192835 bases at least Q30
            Consensus quality: 195619 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 195112; sum-of-contigs
            Quality coverage: 7.73 in Q20 bases; agarose-fp
            Quality coverage: 7.34 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3180: contig of 3180 bp in length
            *    3181     3280: gap of unknown length
            *    3281    12431: contig of 9151 bp in length
            *   12432    12531: gap of unknown length
            *   12532    23140: contig of 10609 bp in length
```

```
            *    23141    23240: gap of unknown length
            *    23241    34935: contig of 11695 bp in length
            *    34936    35035: gap of unknown length
            *    35036    50440: contig of 15405 bp in length
            *    50441    50540: gap of unknown length
            *    50541    71299: contig of 20759 bp in length
            *    71300    71399: gap of unknown length
            *    71400    96045: contig of 24646 bp in length
            *    96046    96145: gap of unknown length
            *    96146   124103: contig of 27958 bp in length
            *   124104   124203: gap of unknown length
            *   124204   160061: contig of 35858 bp in length
            *   160062   160161: gap of unknown length
            *   160162   193415: contig of 33254 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193415
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-480C22"
     misc_feature    1. .3180
                     /note="assembly_name:Contig21"
     misc_feature    3281. .12431
                     /note="assembly_name:Contig22"
     misc_feature    12532. .23140
                     /note="assembly_name:Contig23"
     misc_feature    23241. .34935
                     /note="assembly_name:Contig24"
     misc_feature    35036. .50440
                     /note="assembly_name:Contig25"
     misc_feature    50541. .71299
                     /note="assembly_name:Contig26"
     misc_feature    71400. .96045
                     /note="assembly_name:Contig27"
     misc_feature    96146. .124103
                     /note="assembly_name:Contig28"
     misc_feature    124204. .160061
                     /note="assembly_name:Contig29"
     misc_feature    160162. .193415
                     /note="assembly_name:Contig30
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    57779 a  38131 c  37181 g  59416 t    908 others
ORIGIN


  Query Match             40.4%;  Score 36.4;  DB 2;  Length 193415;
  Best Local Similarity   64.0%;  Pred. No. 57;
  Matches   55;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       3 gaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgg 62
           |||||| ||||||| |||  ||||||   ||| ||| |||| |||||||  |  | ||
Db   21903 GAAAAATATTATTTAAAATATTACTGCTTTTTCTATATTTTTATGTTTTTCTGGTTTTTC 21844

Qy      63 agcaataagtccaatgaaagcaagtg 88
             | | |||     |  |||  | ||
Db   21843 TACCACAAGAATCTTATAAGTCAATG 21818


RESULT  12
CNS01BGA
LOCUS       CNS01BGA      148 bp    mRNA            PLN       02-SEP-1999
DEFINITION  Botrytis cinerea strain T4 cDNA library under conditions of
            nitrogen deprivation.
ACCESSION   AL114194
VERSION     AL114194.1  GI:5828813
KEYWORDS    cDNA library; nitrogen deprivation.
SOURCE      Botryotinia fuckeliana.
  ORGANISM  Botryotinia fuckeliana
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
            Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE   1  (bases 1 to 148)
  AUTHORS   Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
            78026 Versailles, France
REFERENCE   2  (bases 1 to 148)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
            CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The cDNA library to be analyzed within the framework of this
            project was created using a Botrytis cinerea strain which was grown
            under conditions of nitrogen deprivation, which is the normal
            situation for B. cinerea during its development on its host plant.
            The library was produced in an oriented direction, in the pBSII
            vector.
FEATURES             Location/Qualifiers
     source          1. .148
                     /organism="Botryotinia fuckeliana"
                     /strain="T4"
                     /db_xref="taxon:40559"
                     /note="Genoscope sequence ID : W52A031"
BASE COUNT       33 a     17 c     25 g     73 t
ORIGIN


  Query Match             39.3%;  Score 35.4;  DB 8;  Length 148;
  Best Local Similarity   66.2%;  Pred. No. 1e+02;
  Matches   51;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      11 ttatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaataa 70
           ||||||     ||| || | |||||| |  |||  | |||| || || ||||| ||||||
Db      45 TTATTTTTTTTTTTTCTATTATTTTTTTTCTTTTTTTTTTTTTTTTTTTTTGGACCAATAA 104

Qy      71 gtccaatgaaagcaagt 87
             |  ||   | |  ||||
Db     105 GAGCATCAACATTAAGT 121


RESULT  13
AF099924/c
LOCUS       AF099924    33460 bp    DNA             INV       09-AUG-2001
DEFINITION  Caenorhabditis elegans cosmid K07A9, complete sequence.
ACCESSION   AF099924
VERSION     AF099924.2  GI:6671811
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 33460)
  AUTHORS   The C. elegans Sequencing Consortium.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science. 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
REFERENCE   2  (bases 1 to 33460)
  AUTHORS   Davidson,S. and O'Neal,D.
  TITLE     The sequence of C. elegans cosmid K07A9
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 33460)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 33460)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 33460)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

```
REFERENCE   6  (bases 1 to 33460)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 33460)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     On Jan 5, 2000 this sequence version replaced gi:3790764.
            Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

                        NEIGHBORING COSMID INFORMATION

            The 5' clone is Y38C1B, 200 bp overlap;3' clone is W03A6, 200 bp
            overlap.  Actual start of this clone is at base position 1 of
            CELK07A9; actual end is at 11466 of CELW03A6.

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES            Location/Qualifiers
     source         1. .33460
                    /organism="Caenorhabditis elegans"
                    /strain="Bristol N2"
                    /db_xref="taxon:6239"
                    /chromosome="IV"
                    /clone="K07A9"
     gene           96. .9167
                    /gene="K07A9.2"
                    /note="cmk-1"
     CDS            join(96. .200,6063. .6181,7483. .7639,7732. .7997,8434. .8537,
                    8587. .8756,9042. .9167)
                    /gene="K07A9.2"
                    /note="C. elegans Ca2+/calmodulin-dependent protein kinase
                    I (GB:AB021864); contains similarity to Pfam domain
                    PF00069 (pkinase), Score=262.3, E-value=2e-75, N=1; coded
                    for by the following C. elegans cDNAs: yk306d5.3,
                    yk306d5.5, AB021864"
                    /codon_start=1
                    /product="Hypothetical protein K07A9.2"
                    /protein_id="AAF23187.1"
                    /db_xref="GI:6671812"
                    /translation="MPLFKRRDGSGPAPNATIREKYDFRDVLGTGAFSKVFLAESKSD
                    AGQMYAVKCIDKKALKGKEESLENEIKVLRKLRHNNIVQLFDTYDEKQFVYLVMELVT
                    GGELFDRIVAKGSYTEQDASNLIRQVLEAVGFMHDNGVVHRDLKPENLLYYNQDEDSK
                    IMISDFGLSKTEDSGVMATACGTPGYVAPEVLQQKPYGKAVDVWSIGVIAYILLCGYP
                    PFYDESDANLFAQIIKGEYEFDAPYWDQISDSAKDFITHLMCCDPEARFTCQDALSHP
                    WISGNTAYTHDIHGTVAVHLKKSLAKRNWKKAYNAAAAIRQLQMLRLSSNSNRLQKQA
                    SQQQPEPPTPAFHA"
     gene           10199. .11708
                    /gene="K07A9.3"
     CDS            join(10199. .10231,10572. .10658,10709. .10810,11291. .11470,
                    11541. .11708)
                    /gene="K07A9.3"
                    /codon_start=1
                    /product="Hypothetical protein K07A9.3"
                    /protein_id="AAC68811.1"
                    /db_xref="GI:3790766"
                    /translation="MDASTSEFATDFIGIISNEKFSGLDDFGRTDFLRIYDKAKKNKS
                    IKSITRHDVNFRAYILQVLHGKIVVVDRKLLYNVNNYLGFLNIQPMVFAQLEEISIQK
                    KRKKKQTKVPSPPNFRASSLFFQKIGIQNSQFKTCCFEVSIPLIFSCGSLTEDGLVLF
                    TFNIAPYQGNINYRDLMVLALPDEARSLF"
BASE COUNT     9970 a   6275 c   6181 g  11034 t
ORIGIN


  Query Match           39.3%;  Score 35.4;  DB 3;  Length 33460;
  Best Local Similarity  63.5%;  Pred. No. 96;
  Matches   54;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      6 aaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagc 65
          ||||  | ||| ||| ||| | | |||||||||  ||| ||  ||||| | || ||  ||
Db  27539 AAAAGCTTTTTGAAAATTTTCAGAAATTTTTAAATTTCCACATTTTCATTTTTTTTAAGG 27480

Qy     66 aataagtccaatgaaagcaagtgca 90
            | || | |   ||   || || ||
Db  27479 CAGAAATTCTGGGATGCCACGTACA 27455


RESULT  14
AF067612/c
LOCUS       AF067612    69431 bp    DNA             INV       03-JUL-2001
DEFINITION  Caenorhabditis elegans cosmid F55A8, complete sequence.
ACCESSION   AF067612
VERSION     AF067612.2  GI:4982487
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 69431)
  AUTHORS   The C. elegans Sequencing Consortium.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
REFERENCE   2  (bases 1 to 69431)
  AUTHORS   Langston,Y., Wohldmann,P. and Duckels,G.
  TITLE     The sequence of C. elegans cosmid F55A8
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 69431)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 69431)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 69431)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 69431)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     On Jun 3, 1999 this sequence version replaced gi:3193146.
            Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
```

St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 IRQ, England
email: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION

The 5' clone is W03A6, 22600 bp overlap;3' clone is Y48A5B, 2100 bp overlap. Actual start of this clone is at base position 22401 of CELF55A8; actual end is at 67531 of CELF55A8.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers

source 1. .69431
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="F55A8"

gene 104. .31973
/gene="F55A8.2"

CDS join(104. .259,12869. .13036,14716. .14951,16153. .16319, 17407. .17503,23798. .23921,25619. .25726,27685. .28241, 29221. .29854,31878. .31973)
/gene="F55A8.2"
/note="contains similarity to Pfam domains PF00027 (cNMP_binding, Score=246.9, E-value=9.3e-71, N=2), PF00069 (pkinase, Score=250.6, E-value=6.8e-72, N=1) and PF00433 (pkinase_C, Score=16.7, E-value=0.00024, N=1); coded for by the following C. elegans cDNAs: yk2f8.5, yk4g10.3, yk17d12.3, yk30f8.5, yk41a5.3, yk52d1.3, yk52d1.5, yk55a6.3, yk55a6.5, yk55f2.3, yk55f2.5, yk81h11.3, yk81h11.5, yk85a3.5, yk87g1.3, yk243f5.5, yk317d2.5, yk325g6.5, yk328c10.5, yk341c9.5, yk349d8.5, yk385d7.5, yk395h7.5, yk410a4.5, yk443d2.5, yk469c5.5, yk479h11.5, yk486b10.5"
/codon_start=1
/product="Hypothetical protein F55A8.2a"
/protein_id="AAD36954.1"
/db_xref="GI:4982490"
/translation="MSSGSRPSSGGGGGGGGASGGAGGGAPGGGGGGIRGFFSKLRKP SDQPNGNQVQVGTRTFEAHELQKLIPQLEEAISRKDAQLRQQQTIVEGHIKRISELEG EVTTLQRECDKLRSVLEQKAQSAASPGGQPPSPSPRTDQLGNDLQQKAVLPADGVQRA KKIAVSAEPTNFENKPATLQHYNKTVGAKQMIRDAVQKNDFLKQLAKEQIIELVNCMY EMRARAGQWVIQEGEPGDRLFVVAEGELQVSREGALLGKMRAGTWVGELAILYNCTRT ASVQALTDVQLWVLDRSVFQMITQRLGMERHSQLMNFLTKVSIFQNLSEDRISKMADV MDQDYYDGGHYIIRQGEKGDAFFVINSGQVKVTQQIEGETEPREIRVLNQGDFFGERA LLGEEVRTANIIAQAPGVEVLTLDRESFGKLIGDLESLKKDYGDKERLAQVVREPPSP VKIVDDFREEFAQVTLKNVKRLATLGVGGFGRVELVCVNGDKAKTFALKALKKKHIVD TRQQEHIFAERNIMMETSTDWIVKLYKTFRDQKFVYMLLEVCLGGELWTTLRDRGHFD DYTARFYVACVLEGLEYLHRKNIVYRDLKPENCLLANTGYLKLVDFGFAKKLASGRKT WTFCGTPEYVSPEIILNKGHDQAADYWALGIYICELMLGRPPFQASDPMKTYTLILKG VDALEIPNRRRIGKTATALVKKLCRDNPGERLGSGSGGVNDIRKHRWFMGFDWEGLRSR TLKPPILPKVSNPADVTNFDNYPPDNDVPPDEFSGWDEGF"

CDS join(1698. .1724,12869. .13036,14716. .14951,16153. .16319, 17407. .17503,23798. .23921,25619. .25726,27685. .28241, 29221. .29854,31878. .31973)
/gene="F55A8.2"
/note="contains similarity to Pfam domains PF00027 (cNMP_binding, Score=246.9, E-value=9.3e-71, N=2), PF00069 (pkinase, Score=250.6, E-value=6.8e-72, N=1) and PF00433 (pkinase_C, Score=16.7, E-value=0.00024, N=1); coded for by the following C. elegans cDNAs: yk2f8.5, yk4g10.3, yk4g10.5, yk8h2.5, yk17d12.3, yk30f8.5, yk41a5.3, yk41a5.5, yk52d1.3, yk52d1.5, yk55a6.3, yk55a6.5, yk55f2.3, yk55f2.5, yk81h11.3, yk81h11.5, yk87g1.3, yk87g1.5, yk87g12.3, yk87g12.5, yk193d8.5, yk243f5.5, yk285h5.5, yk317d2.5, yk348d12.5, yk349d8.5, yk385d7.5, yk395h7.5, yk410a4.5, yk443d2.5, yk469c5.5, yk486b10.5"
/codon_start=1
/product="Hypothetical protein F55A8.2b"
/protein_id="AAD36953.1"
/db_xref="GI:4982489"
/translation="MKQQPPRIYVQVGTRTFEAHELQKLIPQLEEAISRKDAQLRQQQ TIVEGHIKRISELEGEVTTLQRECDKLRSVLEQKAQSAASPGGQPPSPSPRTDQLGND LQQKAVLPADGVQRAKKIAVSAEPTNFENKPATLQHYNKTVGAKQMIRDAVQKNDFLK QLAKEQIIELVNCMYEMRARAGQWVIQEGEPGDRLFVVAEGELQVSREGALLGKMRAG TVMGELAILYNCTRTASVQALTDVQLWVLDRSVFQMITQRLGMERHSQLMNFLTKVSI FQNLSEDRISKMADVMDQDYYDGGHYIIRQGEKGDAFFVINSGQVKVTQQIEGETEPR EIRVLNQGDFFGERALLGEEVRTANIIAQAPGVEVLTLDRESFGKLIGDLESLKKDYG DKERLAQVVREPPSPVKIVDDFREEFAQVTLKNVKRLATLGVGGFGRVELVCVNGDKA KTFALKALKKKHIVDTRQQEHIFAERNIMMETSTDWIVKLYKTFRDQKFVYMLLEVCL GGELWTTLRDRGHFDDYTARFYVACVLEGLEYLHRKNIVYRDLKPENCLLANTGYLKL VDFGFAKKLASGRKTWTFCGTPEYVSPEIILNKGHDQAADYWALGIYICELMLGRPPF QASDPMKTYTLILKGVDALEIPNRRIGKTATALVKKLCRDNPGERLGSGSGGVNDIRK HRWFMGFDWEGLRSRTLKPPILPKVSNPADVTNFDNYPPDNDVPPDEFSGWDEGF"

gene 65506. .69394
/gene="F55A8.1"

CDS join(65506. .65631,65677. .66207,66728. .66964,69158. .69394)
/gene="F55A8.1"
/note="contains similarity to Pfam domain PF00320 (GATA), Score=42.5, E-value=3e-09, N=1; coded for by the following C. elegans cDNAs: yk96g8.5, yk115f7.5, yk330b8.5, yk330f7.5, yk334d12.5, yk341e9.5, yk345d6.5, yk390a2.5, yk404g2.5, yk446e9.5, yk474a4.5, yk484g12.5, yk524e12.5"
/codon_start=1
/product="Hypothetical protein F55A8.1"
/protein_id="AAD36952.2"
/db_xref="GI:14589827"
/translation="MSISIMTETRPESAEQQHHEVLQRPSDEPCSGCKQLQKDVAKTI SMVMERMDKLQYRLDELLKENNELKSSSVSSGKASPSPAESRSSPKLVETVVAPVSGA RKRKPKERSPPAAASPLPDFSNLMNGFMFDPLNMSNPNGMMQLLSMVQQQQQQQQHHQ HIENQQSVSPPQSKSVKIEDPMDQDVKQEESERSDIPTATEAQNLLDALTAQFSSNGQ ATSTTSPPSSSSQVQAVIEAVATPSSQSQDSSMFEKTETSGDPNAARCSNCRTDKTTA WRRDAEGKLVCNPCGLYYRLHKVRRPIEMRKNHIQQRYRRKNKEKESSAATQIFNQLL TQMPTMATGGVSTDGAINTFNLLEQISQFTQAQELMNSSATF"

BASE COUNT 21037 a 13563 c 12501 g 22330 t
ORIGIN

Query Match 39.3%; Score 35.4; DB 3; Length 69431;
Best Local Similarity 63.5%; Pred. No. 95;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy       6 aaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagc 65
           ||||  | ||| ||| ||| | | ||||||||  ||| ||  |||||  | || ||  ||
Db    4004 AAAAGCTTTTTGAAAATTTTCAGAAATTTTTAAATTTCCACATTTTCATTTTTTTTAAGG 3945

Qy      66 aataagtccaatgaaagcaagtgca 90
             | ||  | |   ||   || || ||
Db    3944 CAGAAATTCTGGGATGCCACGTACA 3920
```

RESULT 15
AC006364/c
LOCUS AC006364 132319 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens BAC clone GS1-146J4 from 7q31.1-31.3, complete sequence.
ACCESSION AC006364
VERSION AC006364.3 GI:4753253
KEYWORDS HTG.

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132319)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 132319)
  AUTHORS   Ryan,E., Bauer,C., Tucci,S. and Spalding,L.
  TITLE     The sequence of Homo sapiens BAC clone GS1-146J4
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 132319)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 132319)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 132319)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 132319)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 5, 1999 this sequence version replaced gi:4337266.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_GS146J04
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone is from the first BAC library from Genome Systems, Inc.
            (http://www.genomesystems.com).
            Cell line:  lymphoblastoid
            Haplotypes:  two
            VECTOR:  pBeloBAC
            Selection:  chloramphenicol
            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of GS1-146J4;
            actual end is at 132319 of GS1-146J4.
FEATURES              Location/Qualifiers
     source           1. .132319
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /map="7q31.1-31.3"
                      /clone="GS1-146J4"
                      /clone_lib="GSBAC1"
     repeat_region    522. .601
                      /rpt_family="L2"
     repeat_region    1216. .1506
                      /rpt_family="Alu"
     repeat_region    1514. .1858
                      /rpt_family="L1"
     repeat_region    2118. .2474
                      /rpt_family="MaLR"
     repeat_region    2512. .2783
                      /rpt_family="MER4-group?"
     repeat_region    3328. .3362
                      /rpt_family="(TGGA)n"
     repeat_region    3810. .3833
                      /rpt_family="POLY_G"
     repeat_region    3880. .4030
                      /rpt_family="MER1_type"
     repeat_region    4108. .4397
                      /rpt_family="Retroviral"
     repeat_region    6413. .6446
                      /rpt_family="(CA)n"
     repeat_region    7359. .7395
                      /rpt_family="AT_rich"
     repeat_region    7396. .7704
                      /rpt_family="MER1_type"
     repeat_region    7930. .7982
                      /rpt_family="Retroviral"
     repeat_region    7983. .8089
                      /rpt_family="(TA)n"
     repeat_region    8111. .8253
                      /rpt_family="Retroviral"
     misc_feature     10366. .10769
                      /note="match to EST R59425 (NID:g830120) yh17c08.r1"
     misc_feature     10401. .10850
                      /note="match to EST H18852 (NID:g885092) ym45f11.r1"
     misc_feature     10754. .11029
                      /note="match to EST AA328782 (NID:g1981047)"
     repeat_region    11459. .11522
                      /rpt_family="L2"
     misc_feature     11660. .12114
                      /note="similar to EST AI220397 (NID:g3802600) qg73h11.x1"
     misc_feature     11665. .11725
                      /note="similar to Mus musculus EST AA986059 (NID:g3167448)
                      uc72e08.y1"
     gene             11713. .44221
                      /gene="WUGSC:H_GS146J04.1"
     CDS              join(11713. .11725,15668. .15772,22543. .22572,23259. .23382,
                      30494. .30552,31735. .31785,34555. .34639,43413. .43539,
                      44132. .44221)
                      /gene="WUGSC:H_GS146J04.1"
                      /note="match to Q92520 (PID:g3334194); H_GS146J04.1"
                      /codon_start=1
                      /product="GS3786"
                      /protein_id="AAD54511.1"
                      /db_xref="GI:5870323"
                      /translation="MRVAGAAKLVVAVAVFLLTFYVISQVFEIKMDASLGNLFARSAL
                      DTAARSTKPPRYKCGISKACPEKHFAFKMASGAANVVGPKICLEDNVLMSGVKNNVGR
                      GINVALANGKTGEVLDTKYFDMWGGDVAPFIEFLKAIQDGTIVLMGTYDDGATKLNDE
                      ARRLIADLGSTSITNLGFRDNWVFCGGKGIKTKSPFEQHIKNNKDTNKYEGWPEVVEM
                      EGCIPQKQD"
     misc_feature     11766. .12114
```

```
                /gene="WUGSC:H_GS146J04.1"
                /note="similar to EST AA897626 (NID:g3034240) oj71h05.s1"
     repeat_region   11963. .12120
                /rpt_family="MER2_type"
     repeat_region   12988. .13025
                /rpt_family="(CA)n"
     repeat_region   13228. .13302
                /rpt_family="L2"
     repeat_region   13737. .13885
                /rpt_family="MIR"
     repeat_region   13926. .13948
                /rpt_family="(TA)n"
     repeat_region   13977. .14293
                /rpt_family="L1"
     repeat_region   14556. .14624
                /rpt_family="L2"
     repeat_region   14831. .14908
                /rpt_family="L1"
     misc_feature    15668. .15772
                /gene="WUGSC:H_GS146J04.1"
                /note="similar to Mus musculus EST AA986059 (NID:g3167448)
                uc72e08.y1"
     repeat_region   16199. .16363
                /rpt_family="Alu"
     repeat_region   16710. .17046
                /rpt_family="MER2_type"
     repeat_region   17337. .17695
                /rpt_family="MaLR"
     repeat_region   17730. .17887
                /rpt_family="MER1_type"
     repeat_region   17984. .18221
                /rpt_family="MER2_type"
     repeat_region   20250. .20288
                /rpt_family="AT_rich"
     repeat_region   20375. .20551
                /rpt_family="L1"
     repeat_region   20660. .20941
                /rpt_family="Alu"
     repeat_region   21910. .22000
                /rpt_family="MER1_type"
     misc_feature    22543. .22572
                /gene="WUGSC:H_GS146J04.1"
                /note="similar to Mus musculus EST AA986059 (NID:g3167448)
                uc72e08.y1"
     repeat_region   23073. .23094
                /rpt_family="AT_rich"
     misc_feature    23259. .23353
                /gene="WUGSC:H_GS146J04.1"

  Query Match             39.3%;  Score 35.4;  DB 9;  Length 132319;
  Best Local Similarity   79.2%;  Pred. No. 94;
  Matches   42;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       3 gaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttctta 55
           ||| || |||||||||| ||  | ||| || |||||| ||  ||||||||||
Db 105826 GAATAATATTATTTCAAGGTCGAATGTTATATTTATGCTTAAATGTTTTCTTA 105774


Search completed: January 24, 2002, 02:27:00
Job time: 4473 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 01:51:28 ; Search time 2272.52 Seconds
(without alignments)
425.572 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	37.8	42.0	561	13	AQ533830	AQ533830 RPCI-11-3
c	2	35.6	39.6	302	10	AV247937	AV247937 AV247937
c	3	34.8	38.7	447	13	B67487	B67487 T25M7TR TAM
c	4	34.8	38.7	482	13	AQ968516	AQ968516 LERJD03TF
c	5	34.8	38.7	488	13	AQ968397	AQ968397 LERJC21TR
	6	34.8	38.7	497	13	AQ958266	AQ958266 LERAW47TR
c	7	34.8	38.7	498	13	B77731	B77731 T28N11TF TA
c	8	34.8	38.7	623	13	B30114	B30114 T26F22TFB T
c	9	34.8	38.7	627	13	B78178	B78178 T31J11TF TA
	10	34.8	38.7	640	13	AQ958844	AQ958844 LEREF16TR
c	11	34.8	38.7	643	13	B98363	B98363 T24C11TFB T
c	12	34.8	38.7	660	13	CNS00WZN	AL094241 Arabidops
	13	34.8	38.7	683	13	AQ968517	AQ968517 LERJD03TR
c	14	34.8	38.7	743	13	AQ958265	AQ958265 LERAW47TF
	15	34.8	38.7	798	13	AQ956484	AQ956484 LERAK26TR
c	16	34.4	38.2	607	13	AZ522516	AZ522516 207PbG02
	17	34.4	38.2	638	13	AZ523727	AZ523727 222PbH02
	18	34.2	38.0	923	13	BH132631	BH132631 ENTOP70TR
	19	34	37.8	871	13	AZ680048	AZ680048 ENTKK43TR
	20	33.8	37.6	867	13	CNS00CX5	AL060052 Drosophil
c	21	33.6	37.3	240	10	AU072517	AU072517 AU072517
c	22	33.6	37.3	240	10	AU072714	AU072714 AU072714
c	23	33.6	37.3	306	11	C24362	C24362 C24362 Dict
c	24	33.6	37.3	376	10	AV760889	AV760889 AV760889
c	25	33.6	37.3	413	13	AQ798320	AQ798320 HS_3160_A
	26	33.6	37.3	552	13	AZ065684	AZ065684 RPCI-23-4
	27	33.6	37.3	676	13	AQ635998	AQ635998 RPCI-11-4
	28	33.6	37.3	913	13	BH132369	BH132369 ENTNH76TR
c	29	33.4	37.1	432	10	AA149033	AA149033 zo03h12.s
c	30	33.2	36.9	463	11	BG603509	BG603509 EST502599
	31	33.2	36.9	526	11	BG603424	BG603424 EST502514
c	32	33.2	36.9	652	11	BG603423	BG603423 EST502513
c	33	33	36.7	1101	13	CNS00K0G	AL076967 Drosophil
c	34	32.8	36.4	420	11	Z38090	Z38090 ATTS4230 Ve
c	35	32.8	36.4	473	13	AQ001587	AQ001587 CIT-HSP-2
c	36	32.8	36.4	589	13	AQ026620	AQ026620 CIT-HSP-2
	37	32.8	36.4	843	13	AZ549410	AZ549410 ENTFG80TF
	38	32.8	36.4	899	13	BH136018	BH136018 ENTOL96TF
	39	32.8	36.4	944	13	AZ543633	AZ543633 ENTDL72TF
	40	32.8	36.4	949	13	CNS018DG	AL109198 Drosophil
c	41	32.8	36.4	958	13	AZ687864	AZ687864 ENTHB28TF
	42	32.6	36.2	510	13	AZ384026	AZ384026 1M0141G19
	43	32.6	36.2	553	13	AQ698845	AQ698845 HS_5557_B
c	44	32.4	36.0	366	13	BH014409	BH014409 TDGBL42TH
	45	32.4	36.0	589	13	AZ247746	AZ247746 RPCI-23-9

ALIGNMENTS

RESULT 1
AQ533830/c
LOCUS AQ533830 561 bp DNA GSS 18-MAY-1999
DEFINITION RPCI-11-379L24.TV RPCI-11 Homo sapiens genomic clone RPCI-11-379L24, DNA sequence.
ACCESSION AQ533830
VERSION AQ533830.1 GI:4845520
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 561)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-379L24.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers

```
     source          1. .561
                     /organism="Homo sapiens"
                     /db_xref="GDB:7645439"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-379L24"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       209 a      88 c      72 g     192 t
ORIGIN


  Query Match             42.0%;  Score 37.8;  DB 13;  Length 561;
  Best Local Similarity   73.8%;  Pred. No. 2.6e+02;
  Matches   48;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       4 aaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgga 63
           |||  |||||||||  | ||||  |||  |||||   |||||  | ||||  ||||||  | ||
Db     428 AAAGGAATTATTTCCATGTTTCTTGTCTTTTTTTCATTTTCTTTTTTTTTTATTGGTTGA 369

Qy      64 gcaat 68
           | |||
Db     368 GAAAT 364


RESULT   2
AV247937/c
LOCUS       AV247937      302 bp    mRNA            EST       04-NOV-1999
DEFINITION  AV247937 RIKEN full-length enriched, 0 day neonate head Mus
            musculus cDNA clone 4833401E05 3' similar to X64411 R.norvegicus
            mRNA for 100 kDa protein, mRNA sequence.
ACCESSION   AV247937
VERSION     AV247937.1  GI:6235396
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 302)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .302
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4833401E05"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I."
BASE COUNT       86 a      58 c      62 g      96 t
ORIGIN


  Query Match             39.6%;  Score 35.6;  DB 10;  Length 302;
  Best Local Similarity   64.6%;  Pred. No. 8.9e+02;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       4 aaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgga 63
           |||||||  ||  |||  |||  ||   |||||  ||  ||  |  ||||  |||   |   |||  ||
Db     185 AAAAAAAATAAATCAGAGTGTAAATTAATTGTTCTGCTGTCATCTTTGGCTGGGGTTAGA 126

Qy      64 gcaataagtccaatgaaagcaa 85
             |  |  |||||      ||||||
Db     125 CCTAGAAGTCGGGGATAAGCAA 104


RESULT   3
B67487/c
LOCUS       B67487        447 bp    DNA             GSS       09-DEC-1997
DEFINITION  T25M7TR TAMU Arabidopsis thaliana genomic clone T25M7, DNA
            sequence.
ACCESSION   B67487
VERSION     B67487.1  GI:2666241
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
```

```
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: rounsley@tigr.org
           Seq primer: M13 Reverse
           Class: BAC ends
           High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T25M7"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                     ; Produced by Rod Wing"
BASE COUNT      187 a     64 c     64 g    132 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 447;
  Best Local Similarity   65.4%;  Pred. No. 1e+03;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     379 AAATAAAAATTTGAAAGTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 320

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     319 TTCTGAAAAAAACGAAAG 302


RESULT   4
AQ968516/c
LOCUS       AQ968516      482 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERJD03TF LERG Arabidopsis thaliana genomic clone LERJD03, DNA
            sequence.
ACCESSION   AQ968516
VERSION     AQ968516.1  GI:6796217
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Similar to A. thaliana chloroplast sequence (GB:AP000423)
            Seq primer: TF
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERJD03"
                     /clone_lib="LERG"
                     /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                     sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT      208 a     64 c     78 g    132 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 482;
  Best Local Similarity   65.4%;  Pred. No. 1e+03;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     217 AAATAAAAATTTGAAAGTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 158

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     157 TTCTGAAAAAAAAGAAAG 140


RESULT   5
AQ968397/c
LOCUS       AQ968397      488 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERJC21TR LERG Arabidopsis thaliana genomic clone LERJC21, DNA
            sequence.
ACCESSION   AQ968397
VERSION     AQ968397.1  GI:6796098
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Similar to A. thaliana chloroplast sequence (GB:AP000423)
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERJC21"
                     /clone_lib="LERG"
                     /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                     sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT      186 a     78 c     83 g    141 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 488;
  Best Local Similarity   65.4%;  Pred. No. 1e+03;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db      95 AAATAAAAATTTGAAAGTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 36

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db      35 TTCTGAAAAAAAAGAAAG 18
```

```
RESULT   6
AQ958266
LOCUS       AQ958266       497 bp    DNA          GSS       28-JAN-2000
DEFINITION  LERAW47TR LERA Arabidopsis thaliana genomic clone LERAW47, DNA
            sequence.
ACCESSION   AQ958266
VERSION     AQ958266.1  GI:6785967
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Similar to A. thaliana chloroplast sequence (GB:AP000423)
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAW47"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT      144 a     82 c     69 g    202 t
ORIGIN


  Query Match           38.7%;  Score 34.8;  DB 13;  Length 497;
  Best Local Similarity   65.4%;  Pred. No. 9.9e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     337 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 396

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     397 TTCTGAAAAAAAAGAAAG 414


RESULT   7
B77731/c
LOCUS       B77731         498 bp    DNA          GSS       16-JAN-1998
DEFINITION  T28N11TF TAMU Arabidopsis thaliana genomic clone T28N11, DNA
            sequence.
ACCESSION   B77731
VERSION     B77731.1  GI:2774370
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T28N11TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T28N11"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                     ; Produced by Rod Wing"
BASE COUNT      210 a     68 c     72 g    148 t
ORIGIN


  Query Match           38.7%;  Score 34.8;  DB 13;  Length 498;
  Best Local Similarity   65.4%;  Pred. No. 9.9e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     379 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 320

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     319 TTCTGAAAAAAAAGAAAG 302


RESULT   8
B30114/c
LOCUS       B30114         623 bp    DNA          GSS       13-OCT-1997
DEFINITION  T26F22TFB TAMU Arabidopsis thaliana genomic clone T26F22, DNA
            sequence.
ACCESSION   B30114
VERSION     B30114.1  GI:2516080
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and
            Venter,J.C.
  TITLE     Use of a BAC End Sequence Database To Identify Minimal Overlaps for
            Arabidopsis Genomic Sequencing
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T26F22TRB
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1. .623
```

```
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T26F22"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                     ; Produced by Rod Wing"
BASE COUNT       261 a     87 c     99 g    176 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 623;
  Best Local Similarity   65.4%;  Pred. No. 8.9e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     378 AAATAAAAATTTGAAAGTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 319

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     318 TTCTGAAAAAAAAGAAAG 301


RESULT   9
B78178/c
LOCUS       B78178         627 bp    DNA             GSS       16-JAN-1998
DEFINITION  T31J11TF TAMU Arabidopsis thaliana genomic clone T31J11, DNA
            sequence.
ACCESSION   B78178
VERSION     B78178.1  GI:2774817
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T31J11TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T31J11"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                     ; Produced by Rod Wing"
BASE COUNT       261 a     87 c    100 g    178 t      1 others
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 627;
  Best Local Similarity   65.4%;  Pred. No. 8.9e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     371 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 312

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     311 TTCTGAAAAAAAAGAAAG 294


RESULT  10
AQ958844
LOCUS       AQ958844       640 bp    DNA             GSS       28-JAN-2000
DEFINITION  LEREF16TR LERE Arabidopsis thaliana genomic clone LEREF16, DNA
            sequence.
ACCESSION   AQ958844
VERSION     AQ958844.1  GI:6786545
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Similar to A. thaliana chloroplast sequence (GB:AP000423)
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LEREF16"
                     /clone_lib="LERE"
                     /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                     sheared to 0.6-0.8 Kbp before ligation."
BASE COUNT       181 a     99 c     83 g    277 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 640;
  Best Local Similarity   65.4%;  Pred. No. 8.8e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     147 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 206

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     207 TTCTGAAAAAAAAGAAAG 224


RESULT  11
B98363/c
LOCUS       B98363         643 bp    DNA             GSS       31-MAR-1998
DEFINITION  T24C11TFB TAMU Arabidopsis thaliana genomic clone T24C11, DNA
            sequence.
ACCESSION   B98363
VERSION     B98363.1  GI:3000442
```

```
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T24C11TF T24C11TRB
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T24C11"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                     ; Produced by Rod Wing"
BASE COUNT      268 a     90 c    103 g    181 t      1 others
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 643;
  Best Local Similarity   65.4%;  Pred. No. 8.8e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     378 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 319

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     318 TTCTGAAAAAAAAGAAAG 301


RESULT  12
CNS00WZN/c
LOCUS       CNS00WZN      660 bp    DNA             GSS       28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence T7 end of BAC T12O2 of
            TAMU library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION   AL094241
VERSION     AL094241.1  GI:5295395
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 660)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="TAMU"
                     /clone="T12O2"
                     /note="end : T7"
BASE COUNT      272 a     93 c    107 g    188 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 660;
  Best Local Similarity   65.4%;  Pred. No. 8.7e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
           ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db     371 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 312

Qy      65 caataagtccaatgaaag 82
             | |    || |||||
Db     311 TTCTGAAAAAAAAGAAAG 294


RESULT  13
AQ968517
LOCUS       AQ968517      683 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERJD03TR LERG Arabidopsis thaliana genomic clone LERJD03, DNA
            sequence.
ACCESSION   AQ968517
VERSION     AQ968517.1  GI:6796218
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Similar to A. thaliana chloroplast sequence (GB:AP000423)
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERJD03"
                     /clone_lib="LERG"
                     /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                     sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT      208 a    111 c     88 g    276 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 683;
  Best Local Similarity   65.4%;  Pred. No. 8.5e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;
```

```
Qy      5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
          ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db    512 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 571

Qy     65 caataagtccaatgaaag 82
            | |    || |||||
Db    572 TTCTGAAAAAAAAGAAAG 589


RESULT  14
AQ958265/c
LOCUS       AQ958265      743 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERAW47TF LERA Arabidopsis thaliana genomic clone LERAW47, DNA
            sequence.
ACCESSION   AQ958265
VERSION     AQ958265.1  GI:6785966
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 743)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Similar to A. thaliana chloroplast sequence (GB:AP000423)
            Seq primer: TF
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .743
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAW47"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT      311 a    107 c    114 g    211 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 743;
  Best Local Similarity   65.4%;  Pred. No. 8.2e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
          ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db    534 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 475

Qy     65 caataagtccaatgaaag 82
            | |    || |||||
Db    474 TTCTGAAAAAAAAGAAAG 457


RESULT  15
AQ956484
LOCUS       AQ956484      798 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERAK26TR LERA Arabidopsis thaliana genomic clone LERAK26, DNA
            sequence.
ACCESSION   AQ956484
VERSION     AQ956484.1  GI:6784185
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Similar to A. thaliana chloroplast sequence (GB:AP000423)
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAK26"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT      231 a    115 c    109 g    343 t
ORIGIN


  Query Match             38.7%;  Score 34.8;  DB 13;  Length 798;
  Best Local Similarity   65.4%;  Pred. No. 7.9e+02;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggag 64
          ||| ||  |||| |||||||    |  || || ||  || || ||||||| ||||||| |
Db    164 AAATAAAAATTTGAAAGTTTTGATTTCTTCTTCTGAATTAATTTTTTCTTCTTGTTGGGG 223

Qy     65 caataagtccaatgaaag 82
            | |    || |||||
Db    224 TTCTGAAAAAAAAGAAAG 241


Search completed: January 24, 2002, 01:51:33
Job time: 2485 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:18:46 ; Search time 90.4 Seconds
(without alignments)
225.476 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	31.2	34.7	3337	1	US-08-072-610-1	Sequence 1, Appli
c	2	31.2	34.7	3337	2	US-08-719-822B-1	Sequence 1, Appli
c	3	31.2	34.7	3337	4	US-09-092-458-1	Sequence 1, Appli
	4	30.6	34.0	87350	3	US-08-781-891-79	Sequence 79, Appl
	5	30	33.3	4140	3	US-08-894-731-2	Sequence 2, Appli
	6	29.6	32.9	1703	4	US-08-378-313-18	Sequence 18, Appl
c	7	28.2	31.3	2334	1	US-08-062-632-4	Sequence 4, Appli
c	8	28	31.1	1013	1	US-07-920-519-30	Sequence 30, Appl
c	9	28	31.1	1013	1	US-08-086-410-23	Sequence 23, Appl
c	10	28	31.1	1013	1	US-08-314-586-30	Sequence 30, Appl
	11	27.8	30.9	1584	1	US-07-667-276A-1	Sequence 1, Appli
c	12	27.4	30.4	1100	4	US-07-861-458C-4	Sequence 4, Appli
	13	27.4	30.4	5613	2	US-08-463-418-1	Sequence 1, Appli
	14	27.2	30.2	602	1	US-08-764-100-8	Sequence 8, Appli
c	15	27.2	30.2	642	1	US-08-764-100-13	Sequence 13, Appl
	16	27.2	30.2	643	1	US-08-764-100-7	Sequence 7, Appli
	17	27.2	30.2	662	4	US-08-998-416-185	Sequence 185, App
c	18	27.2	30.2	663	4	US-08-998-416-191	Sequence 191, App
	19	27.2	30.2	665	4	US-08-998-416-937	Sequence 937, App
	20	27.2	30.2	701	4	US-08-998-416-701	Sequence 701, App
	21	27.2	30.2	711	4	US-08-998-416-786	Sequence 786, App
	22	27.2	30.2	724	4	US-08-998-416-683	Sequence 683, App
	23	27.2	30.2	732	4	US-08-998-416-1036	Sequence 1036, Ap
	24	27.2	30.2	767	4	US-08-998-416-472	Sequence 472, App
	25	27.2	30.2	782	4	US-08-998-416-224	Sequence 224, App
	26	27.2	30.2	827	4	US-08-998-416-535	Sequence 535, App
	27	27.2	30.2	828	4	US-08-998-416-538	Sequence 538, App
	28	27.2	30.2	834	4	US-08-998-416-305	Sequence 305, App
c	29	27.2	30.2	854	4	US-08-998-416-534	Sequence 534, App
c	30	27.2	30.2	860	4	US-08-998-416-287	Sequence 287, App
	31	27.2	30.2	2993	1	US-08-764-100-2	Sequence 2, Appli
c	32	27.2	30.2	2993	1	US-08-764-100-10	Sequence 10, Appl
c	33	27.2	30.2	3000	1	US-08-764-100-9	Sequence 9, Appli
	34	27.2	30.2	3001	1	US-08-764-100-1	Sequence 1, Appli
	35	27.2	30.2	14176	1	US-08-307-499-1	Sequence 1, Appli
c	36	27.2	30.2	14176	1	US-08-307-499-14	Sequence 14, Appl
	37	27.2	30.2	14176	4	US-09-299-268-1	Sequence 1, Appli
c	38	27.2	30.2	14176	4	US-09-299-268-14	Sequence 14, Appl
	39	27	30.0	642	3	US-08-817-926-49	Sequence 49, Appl
	40	27	30.0	839	3	US-08-817-926-50	Sequence 50, Appl
	41	27	30.0	1478	3	US-08-817-926-1	Sequence 1, Appli
	42	27	30.0	1478	4	US-09-297-053-2	Sequence 2, Appli
	43	27	30.0	3562	3	US-08-817-926-19	Sequence 19, Appl
c	44	26.8	29.8	2168	3	US-08-749-522-6	Sequence 6, Appli
	45	26.8	29.8	7083	4	US-09-198-839-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-072-610-1/c
; Sequence 1, Application US/08072610
; Patent No. 5532133
;  GENERAL INFORMATION:
;    APPLICANT:  Barnwell, John
;    TITLE OF INVENTION:  Plasmodium vivax Blood Stage Antigens,
;    TITLE OF INVENTION:  Monoclonal Antibodies, and Diagnostic Assays
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby and Darby
;      STREET:  805 Third Ave.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,610
;      FILING DATE:  19930602
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  5986/07686
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)527-7700
;      TELEFAX:  (212)753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3337 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium vivax
;    IMMEDIATE SOURCE:
;      CLONE:  PvMB3.3.1
US-08-072-610-1


  Query Match                 34.7%;  Score 31.2;  DB 1;  Length 3337;
```

```
  Best Local Similarity   63.2%;  Pred. No. 5.3;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      10 attatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaata 69
           | ||| |     | ||  | | |||||||   ||  | || |||||  ||||| ||||  |
Db     138 AATATATATGTATATATAGCAGTTTTTATAGATTATTTTTATCTTACTGTTGAAGCATGA 79

Qy      70 agtccaatgaaagcaa 85
           | |  ||  |||||||
Db      78 ATTAAAAGAAAAGCAA 63


RESULT   2
US-08-719-822B-1/c
; Sequence 1, Application US/08719822B
; Patent No. 5874527
;  GENERAL INFORMATION:
;    APPLICANT:  Barnwell, John
;    TITLE OF INVENTION:  Plasmodium vivax Blood Stage Antigens
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby and Darby
;      STREET:  805 Third Ave.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,822B
;      FILING DATE:  09/30/96
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  5986/17686US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)527-7700
;      TELEFAX:  (212)753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3337 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium vivax
;    IMMEDIATE SOURCE:
;      CLONE:  PvMB3.3.1
US-08-719-822B-1


  Query Match             34.7%;  Score 31.2;  DB 2;  Length 3337;
  Best Local Similarity   63.2%;  Pred. No. 5.3;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      10 attatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaata 69
           | ||| |     | ||  | | |||||||   ||  | || |||||  ||||| ||||  |
Db     138 AATATATATGTATATATAGCAGTTTTTATAGATTATTTTTATCTTACTGTTGAAGCATGA 79

Qy      70 agtccaatgaaagcaa 85
           | |  ||  |||||||
Db      78 ATTAAAAGAAAAGCAA 63


RESULT   3
US-09-092-458-1/c
; Sequence 1, Application US/09092458
; Patent No. 6231861
;  GENERAL INFORMATION:
;    APPLICANT:  Barnwell, John
;    TITLE OF INVENTION:  Plasmodium vivax Blood Stage Antigens,
;    TITLE OF INVENTION:  Monoclonal Antibodies, and Diagnostic Assays
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby and Darby
;      STREET:  805 Third Ave.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/092,458
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/719,821
;      FILING DATE:  09/30/96
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  5986/17686US3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)527-7700
;      TELEFAX:  (212)753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3337 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium vivax
;    IMMEDIATE SOURCE:
;      CLONE:  PvMB3.3.1
US-09-092-458-1


  Query Match             34.7%;  Score 31.2;  DB 4;  Length 3337;
  Best Local Similarity   63.2%;  Pred. No. 5.3;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      10 attatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaata 69
           | ||| |     | ||  | | |||||||   ||  | || |||||  ||||| ||||  |
Db     138 AATATATATGTATATATAGCAGTTTTTATAGATTATTTTTATCTTACTGTTGAAGCATGA 79

Qy      70 agtccaatgaaagcaa 85
           | |  ||  |||||||
Db      78 ATTAAAAGAAAAGCAA 63


RESULT   4
US-08-781-891-79
; Sequence 79, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
```

```
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  79:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  87350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-79


  Query Match             34.0%;  Score 30.6;  DB 3;  Length 87350;
  Best Local Similarity   62.3%;  Pred. No. 9;
  Matches   48;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           ||| || |   |||||  ||| || |||||  |||| |||    | ||| ||| |  |
Db   76835 ATGCAATAGGGTATTTTGAAGGTTTCTGTATGTTTTCTGTAGAAAAGTTATCTCAAAGGG 76894

Qy      61 ggagcaataagtccaat 77
           |||   |||  |||| |
Db   76895 GGATATATACTTCCATT 76911


RESULT   5
US-08-894-731-2
; Sequence 2, Application US/08894731
; Patent No. 6084089
; GENERAL INFORMATION:
;  APPLICANT: MINE, Toshiki
;  APPLICANT: OHYAMA, Akio
;  APPLICANT: HIYOSHI, Toru
;  APPLICANT: KASAOKA, Keisuke
;  TITLE OF INVENTION: COLD-INDUCIBLE PROMOTER SEQUENCE
;  FILE REFERENCE: 760-234P
;  CURRENT APPLICATION NUMBER: US/08/894,731
;  CURRENT FILING DATE: 1997-10-27
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 4140
;   TYPE: DNA
;   ORGANISM: Solanum tuberosum
US-08-894-731-2


  Query Match             33.3%;  Score 30;  DB 3;  Length 4140;
  Best Local Similarity   59.3%;  Pred. No. 10;
  Matches   51;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 tgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttg 61
           | |||||||||| |  || ||   | | |||||||| ||| | | |||  ||||| |
Db     248 taaaaaaaattagctataaattagatattatttttatttttacctttttagttatttca 307

Qy      62 gagcaataagtccaatgaaagcaagt 87
             |  | |      |||  ||  || |
Db     308 aattatttttaaaaatataattaatt 333


RESULT   6
US-08-378-313-18
; Sequence 18, Application US/08378313
; Patent No. 6207881
;  GENERAL INFORMATION:
;    APPLICANT:  THEOLOGIS, ATHANASIOS
;    APPLICANT:  SATO, TAKAHIDO
;    TITLE OF INVENTION:  CONTROL OF FRUIT RIPENING THROUGH
;    TITLE OF INVENTION:  GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/378,313
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/862,493
;      FILING DATE:  02-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  29190-20002.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 856-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1703 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  11..1489
US-08-378-313-18


  Query Match             32.9%;  Score 29.6;  DB 4;  Length 1703;
  Best Local Similarity   64.7%;  Pred. No. 12;
  Matches   44;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
           || ||||| ||| ||  |||  ||  |   |||  |||  |||  | |||| || || ||
Db    1492 ATTAAAAACATTTTTCAAAATATTCATACCATTCATATAGTTTTTTTTTTTTTTTTTTTT 1551

Qy      61 ggagcaat 68
           ||  ||||
Db    1552 GGGTCAAT 1559
```

```
RESULT   7
US-08-062-632-4/c
; Sequence 4, Application US/08062632
; Patent No. 5712090
;  GENERAL INFORMATION:
;    APPLICANT:  Artiushin, Sergey
;    APPLICANT:  Stipkovits, Laslo
;    APPLICANT:  Minion, F. Chris
;    TITLE OF INVENTION:  PCR-Based Assay For Mycoplasma
;    TITLE OF INVENTION:  Hyopneumoniae
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dickstein, Shapiro and Morin
;      STREET:  2101 L. St. NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  22037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/062,632
;      FILING DATE:  18-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brady Jr., James W.
;      REGISTRATION NUMBER:  32,115
;      REFERENCE/DOCKET NUMBER:  I8900.018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)785-9700
;      TELEFAX:  (202)887-0689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2334 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-062-632-4


  Query Match             31.3%;  Score 28.2;  DB 1;  Length 2334;
  Best Local Similarity   64.6%;  Pred. No. 28;
  Matches   42;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
          || || |||| |   | ||   ||  ||| |||||||| ||| | ||| || | || |||
Db    236 ATTAAGAAAAATCGCTTAATTATTGGTGTTATTTTTATTTTTACTTGTGTTTTAATAGTT 177

Qy     61 ggagc 65
             ||
Db    176 TCTGC 172


RESULT   8
US-07-920-519-30/c
; Sequence 30, Application US/07920519
; Patent No. 5382518
;  GENERAL INFORMATION:
;    APPLICANT:  CAPUT, DANIEL
;    APPLICANT:  FERRARA, PASCUAL
;    APPLICANT:  GUILLEMOT, JEAN-CLAUDE
;    APPLICANT:  KAGHAD, MOURAD
;    APPLICANT:  LEGOUX, RICHARD
;    APPLICANT:  LOISON, GERARD
;    APPLICANT:  LARBRE, ELIZABETH
;    APPLICANT:  LUPKER, JOHANNES
;    APPLICANT:  LEPLATOIS, PASCUAL
;    APPLICANT:  SALOME, MARK
;    TITLE OF INVENTION:  URATE OXIDASE ACTIVITY PROTEIN,
;    TITLE OF INVENTION:  RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
;    TITLE OF INVENTION:  MICRO-ORGANISMS AND TRANSFORMED CELLS
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/920,519
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/659,408
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  16781/276 BEDL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1013 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  Fragment D
US-07-920-519-30


  Query Match             31.1%;  Score 28;  DB 1;  Length 1013;
  Best Local Similarity   60.5%;  Pred. No. 30;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     11 ttatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaataa 70
          || |||     |||  | |  ||||| | ||||  | |||| || || |||||   | ||
Db   1004 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAAGTAAAA 945

Qy     71 gtccaatgaaagcaag 86
             | | |||| | | |
Db    944 AACAAGTGAATGAATG 929


RESULT   9
US-08-086-410-23/c
; Sequence 23, Application US/08086410
; Patent No. 5407822
;  GENERAL INFORMATION:
;    APPLICANT:  LEPLATOIS, Pascal
;    APPLICANT:  LOISON, Gerard
;    APPLICANT:  PESSEGUE, Bernard
;    APPLICANT:  SHIRE, David
;    TITLE OF INVENTION:  Artificial promoter for the expression
;    TITLE OF INVENTION:  of proteins in yeast
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  King Street Station,Suite 500,1800 Diagonal
;      STREET:  Road,PO Box 299
```

```
;       CITY:  ALEXANDRIA
;       STATE:  VIRGINIA
;       COUNTRY:  USA
;       ZIP:  22313-0299
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/086,410
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/768,083
;       FILING DATE:  02-OCT-1991
;       APPLICATION NUMBER:  FR 89 17467
;       FILING DATE:  29-DEC-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SAXE, Bernhard D
;       REGISTRATION NUMBER:  28,665
;       REFERENCE/DOCKET NUMBER:  16781/318
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-9300
;       TELEFAX:  (703) 683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  23:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1013 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     IMMEDIATE SOURCE:
;       CLONE:  AccI-BamHI fragment
US-08-086-410-23


  Query Match             31.1%;  Score 28;  DB 1;  Length 1013;
  Best Local Similarity   60.5%;  Pred. No. 30;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      11 ttatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaataa 70
           || |||     |||  | |  ||||| | ||||  | |||| || || |||||   | ||
Db    1004 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAAGTAAAA 945

Qy      71 gtccaatgaaagcaag 86
             | | |||| | | |
Db     944 AACAAGTGAATGAATG 929


RESULT  10
US-08-314-586-30/c
; Sequence 30, Application US/08314586
; Patent No. 5541098
;  GENERAL INFORMATION:
;     APPLICANT:  CAPUT, DANIEL
;     APPLICANT:  FERRARA, PASCUAL
;     APPLICANT:  GUILLEMOT, JEAN-CLAUDE
;     APPLICANT:  KAGHAD, MOURAD
;     APPLICANT:  LEGOUX, RICHARD
;     APPLICANT:  LOISON, GERARD
;     APPLICANT:  LARBRE, ELIZABETH
;     APPLICANT:  LUPKER, JOHANNES
;     APPLICANT:  LEPLATOIS, PASCUAL
;     APPLICANT:  SALOME, MARK
;     APPLICANT:  LAURENT, PATRICK
;     TITLE OF INVENTION:  URATE OXIDASE ACTIVITY PROTEIN,
;     TITLE OF INVENTION:  RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
;     TITLE OF INVENTION:  MICROORGANISMS AND TRANSFORMED CELLS
;     NUMBER OF SEQUENCES:  40
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, Suite 500
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/314,586
;       FILING DATE:  28-SEP-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/659,408
;       FILING DATE:  25-APR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  16781/509/BEDL
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)836-9300
;       TELEFAX:  (703)683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1013 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     IMMEDIATE SOURCE:
;       CLONE:  Fragment D
US-08-314-586-30


  Query Match             31.1%;  Score 28;  DB 1;  Length 1013;
  Best Local Similarity   60.5%;  Pred. No. 30;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      11 ttatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaataa 70
           || |||     |||  | |  ||||| | ||||  | |||| || || |||||   | ||
Db    1004 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAAGTAAAA 945

Qy      71 gtccaatgaaagcaag 86
             | | |||| | | |
Db     944 AACAAGTGAATGAATG 929


RESULT  11
US-07-667-276A-1
; Sequence 1, Application US/07667276A
; Patent No. 5470971
;  GENERAL INFORMATION:
;     APPLICANT:  Kondo, Keiji
;     APPLICANT:  Inouye, Masayori
;     TITLE OF INVENTION:  STRESS-INDUCED PROTEINS, GENES CODING
;     TITLE OF INVENTION:  THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
;     TITLE OF INVENTION:  APPLICATIONS
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Weiser & Associates
;       STREET:  230 S. Fifteenth Street, Suite 500
;       CITY:  Philadelphia
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19102
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/667,276A
;      FILING DATE:  11-MAR-1991
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weiser, Gerard J.
;      REGISTRATION NUMBER:  19,763
;      REFERENCE/DOCKET NUMBER:  377.5351P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-875-8383
;      TELEFAX:  215-875-8394
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1584 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
;      STRAIN:  S288C
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  475..1104
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1
;      OTHER INFORMATION:  /note= "Base #1 of Sequence No. 5470971 1
;      OTHER INFORMATION:  corresponds to base -474 of the sequence listed in
;      OTHER INFORMATION:  Figure 4 of the application"
US-07-667-276A-1


  Query Match             30.9%;  Score 27.8;  DB 1;  Length 1584;
  Best Local Similarity   62.0%;  Pred. No. 34;
  Matches   44;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       2 tgaaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttg 61
           | | || || ||||| |||  ||||    ||| ||||   |||   || | | ||||||
Db    1119 TTACAACAAATATTTGAAAAATTACCTCCATTATTATACCTTCTCTTTATGTAATTGTTA 1178

Qy      62 gagcaataagt 72
           |  | | || |
Db    1179 GTTCGAAAATT 1189


RESULT  12
US-07-861-458C-4/c
; Sequence 4, Application US/07861458C
; Patent No. 6232061
;  GENERAL INFORMATION:
;    APPLICANT:  Marchionni, Mark Andrew
;    APPLICANT:  Johnson, Carl D.
;    TITLE OF INVENTION:  HOMOLOGY CLONING
;    NUMBER OF SEQUENCES:  142
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/861,458C
;      FILING DATE:  04/01/92
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  04585/014001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1100
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-861-458C-4


  Query Match             30.4%;  Score 27.4;  DB 4;  Length 1100;
  Best Local Similarity   59.7%;  Pred. No. 42;
  Matches   46;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      11 ttatttcaaagtttactgtaatttttatgttttcatgttttcttattgttggagcaataa 70
           || |||     |||  | |  ||||| | ||||  | |||| || || ||  | ||| ||
Db    1079 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATCAAAAA 1020

Qy      71 gtccaatgaaagcaagt 87
           |  || |  |  ||| |
Db    1019 GCTCACTATATACAATT 1003


RESULT  13
US-08-463-418-1
; Sequence 1, Application US/08463418
; Patent No. 5908971
;  GENERAL INFORMATION:
;    APPLICANT:  Van Der Straeten, Dominique et al.
;    TITLE OF INVENTION:  CRUCIFER ACC SYNTHASE AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,418
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/962,481
;      FILING DATE:  15-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/161002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5613 base pairs
```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-463-418-1

Query Match 30.4%; Score 27.4; DB 2; Length 5613;
Best Local Similarity 65.6%; Pred. No. 47;
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 4 aaaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgttgga 63
Db 2033 AAACAAATTCAGTCGTGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTAAA 2092

Qy 64 g 64
Db 2093 G 2093

RESULT 14
US-08-764-100-8
; Sequence 8, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; TITLE OF INVENTION: Compounds
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-764-100-8

Query Match 30.2%; Score 27.2; DB 1; Length 602;
Best Local Similarity 67.9%; Pred. No. 46;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
Db 535 AATTAATTATTTTATTTTTTATTTTATTTTTTATTTTTATTATTTTTATTTTTATT 590

RESULT 15
US-08-764-100-13/c
; Sequence 13, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; TITLE OF INVENTION: Compounds
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 642 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-764-100-13

Query Match 30.2%; Score 27.2; DB 1; Length 642;
Best Local Similarity 67.9%; Pred. No. 46;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 5 aaaaaattatttcaaagtttactgtaatttttatgttttcatgttttcttattgtt 60
Db 106 AATTAATTATTTTATTTTTTATTTTATTTTTTATTTTTATTATTTTTATTTTTATT 51

Search completed: January 24, 2002, 02:18:58
Job time: 3931 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 03:24:24 ; Search time 93.51 Seconds
(without alignments)
217.977 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 495388

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	13	14.4	24	3	US-08-672-215-1	Sequence 1, Appli
	2	13	14.4	30	2	US-08-629-001A-79	Sequence 79, Appl
	3	13	14.4	30	4	US-08-642-274D-158	Sequence 158, App
c	4	13	14.4	38	5	PCT-US96-00547-40	Sequence 40, Appl
c	5	13	14.4	39	1	US-08-105-483-168	Sequence 168, App
c	6	13	14.4	39	1	US-08-709-209-168	Sequence 168, App
c	7	13	14.4	39	1	US-08-303-275-56	Sequence 56, Appl
c	8	13	14.4	39	1	US-08-458-101-168	Sequence 168, App
	9	12	13.3	14	1	US-08-271-880A-201	Sequence 201, App
	10	12	13.3	14	2	US-08-910-408-201	Sequence 201, App
	11	12	13.3	14	3	US-09-249-215-201	Sequence 201, App
	12	12	13.3	18	3	US-09-163-162-19	Sequence 19, Appl
	13	12	13.3	18	4	US-09-286-407-19	Sequence 19, Appl
c	14	12	13.3	21	1	US-08-661-507-6	Sequence 6, Appli
c	15	12	13.3	21	2	US-08-855-085-4	Sequence 4, Appli
c	16	12	13.3	21	2	US-09-186-030-4	Sequence 4, Appli
c	17	12	13.3	21	2	US-08-865-675-4	Sequence 4, Appli
c	18	12	13.3	21	2	US-08-933-749-5	Sequence 5, Appli
c	19	12	13.3	21	2	US-09-237-510-4	Sequence 4, Appli
c	20	12	13.3	21	3	US-09-120-916-4	Sequence 4, Appli
c	21	12	13.3	21	3	US-08-964-020-8	Sequence 8, Appli
c	22	12	13.3	21	3	US-09-235-583-5	Sequence 5, Appli
c	23	12	13.3	21	4	US-09-599-164-5	Sequence 5, Appli
c	24	12	13.3	23	1	US-08-727-003A-10	Sequence 10, Appl
c	25	12	13.3	23	3	US-08-487-799-11	Sequence 11, Appl
	26	12	13.3	24	2	US-08-210-762E-31	Sequence 31, Appl
	27	12	13.3	24	4	US-08-256-799-26	Sequence 26, Appl
	28	12	13.3	24	4	US-08-462-437-26	Sequence 26, Appl
c	29	12	13.3	25	1	US-08-271-880A-140	Sequence 140, App
c	30	12	13.3	25	2	US-08-910-408-140	Sequence 140, App
c	31	12	13.3	25	3	US-09-249-215-140	Sequence 140, App
	32	12	13.3	26	3	US-08-835-728D-72	Sequence 72, Appl
c	33	12	13.3	26	3	US-08-835-728D-176	Sequence 176, App
	34	12	13.3	26	4	US-09-490-558-72	Sequence 72, Appl
c	35	12	13.3	26	4	US-09-490-558-176	Sequence 176, App
c	36	12	13.3	27	2	US-08-668-128B-6	Sequence 6, Appli
c	37	12	13.3	27	2	US-08-905-445-6	Sequence 6, Appli
c	38	12	13.3	27	3	US-08-644-116A-6	Sequence 6, Appli
	39	12	13.3	28	2	US-08-771-602D-20	Sequence 20, Appl
c	40	12	13.3	28	2	US-08-771-602D-21	Sequence 21, Appl
c	41	12	13.3	30	1	US-08-619-724-3	Sequence 3, Appli
c	42	12	13.3	30	2	US-08-629-001A-25	Sequence 25, Appl
c	43	12	13.3	30	4	US-08-642-274D-104	Sequence 104, App
c	44	12	13.3	30	4	US-08-935-312-14	Sequence 14, Appl
c	45	12	13.3	33	1	US-08-341-456A-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT 1
US-08-672-215-1/c
; Sequence 1, Application US/08672215
; Patent No. 6020121
;  GENERAL INFORMATION:
;    APPLICANT:  Ying Bao, Amy Boggs, Pamela R. Contag,
;    APPLICANT:  Nancy A. Federspiel, Alan Herbert,
;    APPLICANT:  Scott J. Hecker, Francois Malouin
;    TITLE OF INVENTION:  INHIBITORS OF REGULATORY PATHWAYS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/672,215
;      FILING DATE:  June 25, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/004,626
;      FILING DATE:  September 29, 1995
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  219/158
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-672-215-1
```

```
  Query Match            14.4%;  Score 13;  DB 3;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 7e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      27 tgtaatttttatg 39
           |||||||||||||
Db      20 TGTAATTTTTATG 8


RESULT   2
US-08-629-001A-79
; Sequence 79, Application US/08629001A
; Patent No. 5858661
;  GENERAL INFORMATION:
;    APPLICANT:  Shiloh, Yosef
;    TITLE OF INVENTION:  ATAXIA-TELANGIECTASIA GENE AND ITS
;    TITLE OF INVENTION:  GENOMIC ORGANIZATION
;    NUMBER OF SEQUENCES:  139
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 5858661thwestern Hwy.
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,001A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohn, Kenneth I.
;      REGISTRATION NUMBER:  30,955
;      REFERENCE/DOCKET NUMBER:  2290.00032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 539-5050
;      TELEFAX:  (810) 539-5055
;  INFORMATION FOR SEQ ID NO:  79:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-629-001A-79


  Query Match            14.4%;  Score 13;  DB 2;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttc 17
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT   3
US-08-642-274D-158
; Sequence 158, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
;  APPLICANT:  Shiloh, Yosef
;  TITLE OF INVENTION:  MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
;  TITLE OF INVENTION:  SCREEN FOR A PARTIAL A-T PHENOTYPE
;  FILE REFERENCE:  229000033
;  CURRENT APPLICATION NUMBER: US/08/642,274D
;  CURRENT FILING DATE:  1996-05-03
;  NUMBER OF SEQ ID NOS:  220
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 158
;  LENGTH: 30
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:intronic
;  OTHER INFORMATION: sequence
US-08-642-274D-158


  Query Match            14.4%;  Score 13;  DB 4;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttc 17
           |||||||||||||
Db      15 aaaaaattatttc 27


RESULT   4
PCT-US96-00547-40/c
; Sequence 40, Application PC/TUS9600547
;  GENERAL INFORMATION:
;    APPLICANT:  Virogenetics Corporation
;    TITLE OF INVENTION:  RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00547
;      FILING DATE:  12-JAN-1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/372,664
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  38 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US96-00547-40


  Query Match            14.4%;  Score 13;  DB 5;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 6.8e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT   5
US-08-105-483-168/c
```

```
; Sequence 168, Application US/08105483
; Patent No. 5494807
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  462
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/105,483
;      FILING DATE:  12-AUG-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/847,951
;      FILING DATE:  06-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  168:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-105-483-168


  Query Match             14.4%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 6.7e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT   6
US-08-709-209-168/c
; Sequence 168, Application US/08709209
; Patent No. 5762938
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  462
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/709,209
;      FILING DATE:  21-AUG-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/105,483
;      FILING DATE:  12-AUG-1993
;      APPLICATION NUMBER:  US 07/847,951
;      FILING DATE:  06-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  168:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-709-209-168


  Query Match             14.4%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 6.7e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT   7
US-08-303-275-56/c
; Sequence 56, Application US/08303275
; Patent No. 5766598
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  IMMUNODEFICIENCY VIRUS RECOMBINANT
;    TITLE OF INVENTION:  POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  205
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/303,275
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/897,382
;      FILING DATE:  11-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
```

```
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-303-275-56


  Query Match             14.4%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 6.7e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT   8
US-08-458-101-168/c
; Sequence 168, Application US/08458101
; Patent No. 5766599
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Perkus, Marion E.
;    APPLICANT:  Taylor, Jill
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  No. 5766599ton, Elizabeth K.
;    APPLICANT:  Riviere, Michel
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Limbach, Keith J.
;    APPLICANT:  Johnson, Gerard P.
;    APPLICANT:  Pincus, Steven E.
;    APPLICANT:  Cox, William I.
;    APPLICANT:  Audonnet, Jean-Christophe Francis
;    APPLICANT:  Gettig, Russell Robert
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  467
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,101
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2740
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  168:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-458-101-168


  Query Match             14.4%;  Score 13;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 6.7e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT   9
US-08-271-880A-201
; Sequence 201, Application US/08271880A
; Patent No. 5693535
;  GENERAL INFORMATION:
;    APPLICANT:  Kenneth G. Draper
;    APPLICANT:  Bharat Chowrira
;    APPLICANT:  James McSwiggen
;    APPLICANT:  Dan T. Stinchcomb
;    APPLICANT:  James D. Thompson
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR INHIBITING
;    TITLE OF INVENTION:  HUMAN IMMUNODEFICIENCY VIRUS
;    TITLE OF INVENTION:  REPLICATION
;    NUMBER OF SEQUENCES:  232
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/271,880A
;      FILING DATE:  July 7, 1994
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including appplication
;    PRIOR APPLICATION DATA:  described below:             two
;      APPLICATION NUMBER:  08/103,243
;      FILING DATE:  August 6, 1993
;      APPLICATION NUMBER:  07/882,886
;      FILING DATE:  May 14, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  206/116
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  201:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-271-880A-201


  Query Match             13.3%;  Score 12;  DB 1;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 CAATGAAAGCAA 13


RESULT  10
```

```
US-08-910-408-201
; Sequence 201, Application US/08910408
; Patent No. 5972704
;  GENERAL INFORMATION:
;    APPLICANT:  Kenneth G. Draper
;    APPLICANT:  Bharat Chowrira
;    APPLICANT:  James McSwiggen
;    APPLICANT:  Dan T. Stinchcomb
;    APPLICANT:  James D. Thompson
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR INHIBITING
;    TITLE OF INVENTION:  HUMAN IMMUNODEFICIENCY VIRUS
;    TITLE OF INVENTION:  REPLICATION
;    NUMBER OF SEQUENCES:  232
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,408
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/271,880
;      FILING DATE:  July 7, 1994
;      APPLICATION NUMBER:  08/103,243
;      FILING DATE:  August 6, 1993
;      APPLICATION NUMBER:  07/882,886
;      FILING DATE:  May 14, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  206/116
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  201:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-910-408-201


  Query Match             13.3%;  Score 12;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 CAATGAAAGCAA 13
```

```
RESULT  11
US-09-249-215-201
; Sequence 201, Application US/09249215
; Patent No. 6159692
;   GENERAL INFORMATION:
;        APPLICANT: Kenneth G. Draper
;                   Bharat Chowrira
;                   James McSwiggen
;                   Dan T. Stinchcomb
;                   James D. Thompson
;        TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
;                            HUMAN IMMUNODEFICIENCY VIRUS
;                            REPLICATION
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Lyon & Lyon
;             STREET: 633 West Fifth Street
;                     Suite 4700
;             CITY: Los Angeles
;             STATE: California
;             COUNTRY: U.S.A.
;             ZIP: 90071
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;                          storage
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: IBM P.C. DOS 5.0
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/249,215
;             FILING DATE: 12-Feb-1999
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/910,408
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/103,243
;             FILING DATE: August 6, 1993
;             APPLICATION NUMBER: 07/882,886
;             FILING DATE: May 14, 1992
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Warburg, Richard
;             REGISTRATION NUMBER: 32,327
;             REFERENCE/DOCKET NUMBER: 206/116
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (213) 489-1600
;             TELEFAX: (213) 955-0440
;             TELEX: 67-3510
;   INFORMATION FOR SEQ ID NO: 201:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 201:
US-09-249-215-201


  Query Match             13.3%;  Score 12;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 CAATGAAAGCAA 13
```

```
RESULT  12
US-09-163-162-19
; Sequence 19, Application US/09163162
; Patent No. 6077709
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Ackermann, Elizabeth J.
;  APPLICANT: Swayze, Eric E.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF Survivin EXPRESSION
;  FILE REFERENCE: RTS-0008
;  CURRENT APPLICATION NUMBER: US/09/163,162
;  CURRENT FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 19
;   LENGTH: 18
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
```

```
;   FEATURE:
;   OTHER INFORMATION: Antisense Oligonucleotide
US-09-163-162-19


  Query Match             13.3%;  Score 12;  DB 3;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      51 tcttattgttgg 62
           ||||||||||||
Db       1 tcttattgttgg 12


RESULT  13
US-09-286-407-19
; Sequence 19, Application US/09286407A
; Patent No. 6165788
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Ackermann, Elizabeth J.
;  APPLICANT: Swayze, Eric E.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF Survivin EXPRESSION
;  FILE REFERENCE: ISPH-0349
;  CURRENT APPLICATION NUMBER: US/09/286,407A
;  CURRENT FILING DATE: 1999-04-05
;  NUMBER OF SEQ ID NOS: 48
; SEQ ID NO 19
;   LENGTH: 18
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Antisense Oligonucleotide
US-09-286-407-19


  Query Match             13.3%;  Score 12;  DB 4;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      51 tcttattgttgg 62
           ||||||||||||
Db       1 tcttattgttgg 12


RESULT  14
US-08-661-507-6/c
; Sequence 6, Application US/08661507
; Patent No. 5814490
;  GENERAL INFORMATION:
;    APPLICANT:  Spears, Patricia A.
;    TITLE OF INVENTION:  AMPLIFICATION AND DETECTION OF CHLAMYDIA
;    TITLE OF INVENTION:  TRACHOMATIS NUCLEIC ACIDS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/661,507
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3489
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-661-507-6


  Query Match             13.3%;  Score 12;  DB 1;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  15
US-08-855-085-4/c
; Sequence 4, Application US/08855085
; Patent No. 5846726
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Schram, James L.
;    APPLICANT:  Linn, Carl P.
;    APPLICANT:  Vonk, Glenn P.
;    APPLICANT:  Walker, George T.
;    TITLE OF INVENTION:  Detection of Nucleic Acids by
;    TITLE OF INVENTION:  Fluorescence Quenching
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,085
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3747
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-855-085-4


  Query Match             13.3%;  Score 12;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  16
```

```
US-09-186-030-4/c
; Sequence 4, Application US/09186030
; Patent No. 5919630
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Schram, James L.
;    APPLICANT:  Linn, Carl P.
;    APPLICANT:  Vonk, Glenn P.
;    APPLICANT:  Walker, George T.
;    TITLE OF INVENTION:  Detection of Nucleic Acids by
;    TITLE OF INVENTION:  Fluorescence Quenching
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/186,030
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/855,085
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3747
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-186-030-4


  Query Match             13.3%;  Score 12;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  17
US-08-865-675-4/c
; Sequence 4, Application US/08865675
; Patent No. 5928869
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Linn, Carl P.
;    APPLICANT:  Schram, James L.
;    TITLE OF INVENTION:  DETECTION OF NUCLEIC ACIDS BY
;    TITLE OF INVENTION:  FLUORESCENCE QUENCHING
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/865,675
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3746
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-865-675-4


  Query Match             13.3%;  Score 12;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  18
US-08-933-749-5/c
; Sequence 5, Application US/08933749
; Patent No. 5935791
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Hsieh, Helen V.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Linn, Carl P.
;    TITLE OF INVENTION:  Detection of Nucleic Acids by
;    TITLE OF INVENTION:  Fluorescence Quenching
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,749
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3749
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-933-749-5


  Query Match             13.3%;  Score 12;  DB 2;  Length 21;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  19
US-09-237-510-4/c
; Sequence 4, Application US/09237510
; Patent No. 5958700
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Linn, Carl P.
;    APPLICANT:  Schram, James L.
;    TITLE OF INVENTION:  DETECTION OF NUCLEIC ACIDS BY
;    TITLE OF INVENTION:  FLUORESCENCE QUENCHING
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/237,510
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3746
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-237-510-4


  Query Match             13.3%;  Score 12;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  20
US-09-120-916-4/c
; Sequence 4, Application US/09120916
; Patent No. 6054279
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Schram, James L.
;    APPLICANT:  Linn, Carl P.
;    APPLICANT:  Vonk, Glenn P.
;    APPLICANT:  Walker, George T.
;    TITLE OF INVENTION:  Detection of Nucleic Acids by
;    TITLE OF INVENTION:  Fluorescence Quenching
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,916
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/855,085
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3747
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-120-916-4


  Query Match             13.3%;  Score 12;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  21
US-08-964-020-8/c
; Sequence 8, Application US/08964020
; Patent No. 6077669
;  GENERAL INFORMATION:
;    APPLICANT:  Vonk, Glenn P.
;    APPLICANT:  Little, Michael C.
;    TITLE OF INVENTION:  Kit and Method for Fluorescence Based
;    TITLE OF INVENTION:  Detection Assay
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick - Becton, Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/964,020
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highet, David W.
;      REGISTRATION NUMBER:  30,265
;      REFERENCE/DOCKET NUMBER:  p-4025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 847-5317
```

```
;      TELEFAX:  (201) 848-9228
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  21 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-964-020-8


  Query Match             13.3%;  Score 12;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  22
US-09-235-583-5/c
; Sequence 5, Application US/09235583
; Patent No. 6130047
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Hsieh, Helen V.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Linn, Carl P.
;    TITLE OF INVENTION:  Detection of Nucleic Acids by
;    TITLE OF INVENTION:  Fluorescence Quenching
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/235,583
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3749
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-235-583-5


  Query Match             13.3%;  Score 12;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  23
US-09-599-164-5/c
; Sequence 5, Application US/09599164
; Patent No. 6261784
;  GENERAL INFORMATION:
;    APPLICANT:  Nadeau, James G.
;    APPLICANT:  Hsieh, Helen V.
;    APPLICANT:  Pitner, James B.
;    APPLICANT:  Linn, Carl P.
;    TITLE OF INVENTION:  Detection of Nucleic Acids by
;    TITLE OF INVENTION:  Fluorescence Quenching
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  R. J. Rodrick, Becton Dickinson and Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/599,164
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,749
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3749
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-599-164-5


  Query Match             13.3%;  Score 12;  DB 4;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  24
US-08-727-003A-10/c
; Sequence 10, Application US/08727003A
; Patent No. 5804383
;  GENERAL INFORMATION:
;    APPLICANT:  Gruenert, Dieter, C.
;    APPLICANT:  Dohrman, Austin F.
;    TITLE OF INVENTION:  A METHOD AND ASSAY FOR
;    TITLE OF INVENTION:  DETECTIONOF THE EXPRESSION
;    TITLE OF INVENTION:  OF ALLELE-SPECIFIC MUTATIONS
;    TITLE OF INVENTION:  BY ALLELE-SPECIFIC IN SITU
;    TITLE OF INVENTION:  REVERSE TRANSCRIPTASE
;    TITLE OF INVENTION:  POLYMERASE CHAIN REACTION
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIK A, L.L.P.
;      STREET:  385 Sherman Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Kb storage
;      COMPUTER:  PC
```

```
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/727,003A
;      FILING DATE:  October 8, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/005,254
;      FILING DATE:  October 10, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hana Verny
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480-77
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)324-1677
;      TELEFAX:  (415)324-1678
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  synthetic oligonucleotide
US-08-727-003A-10


  Query Match             13.3%;  Score 12;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      59 ttggagcaataa 70
           ||||||||||||
Db      23 TTGGAGCAATAA 12


RESULT  25
US-08-487-799-11/c
; Sequence 11, Application US/08487799C
; Patent No. 6010908
; GENERAL INFORMATION:
;  APPLICANT: Gruenert, Deiter C.
;  APPLICANT: Kunzelmann, Karl
;  TITLE OF INVENTION: GENE THERAPY BY SMALL FRAGMENTS HOMOLOGOUS REPLACEMENT
;  FILE REFERENCE: 480.18-1(HV)
;  CURRENT APPLICATION NUMBER: US/08/487,799C
;  CURRENT FILING DATE: 1995-06-07
;  EARLIER APPLICATION NUMBER: 07/933,471
;  EARLIER FILING DATE: 1992-08-21
;  EARLIER APPLICATION NUMBER: 08/409,544
;  EARLIER FILING DATE: 1995-03-24
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 23
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: oligonucleotide
US-08-487-799-11


  Query Match             13.3%;  Score 12;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      59 ttggagcaataa 70
           ||||||||||||
Db      23 TTGGAGCAATAA 12


RESULT  26
US-08-210-762E-31
; Sequence 31, Application US/08210762E
; Patent No. 5837441
;  GENERAL INFORMATION:
;    APPLICANT:  Hjelle, Brian
;    APPLICANT:  Jenison, Steve
;    TITLE OF INVENTION:  Molecular Clones Producing Recombinant DNA Antigens of
;    TITLE OF INVENTION:  the HARDS Virus.
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffman, Wasson & Gitler
;      STREET:  2361 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 500 Kb storage
;      COMPUTER:  Accel 486
;      OPERATING SYSTEM:  Windows 3.1
;      SOFTWARE:  WordPerfect 6.1 for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/210,762E
;      FILING DATE:  22-MAR-94
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/141,035
;      FILING DATE:  26-OCT-93
;      APPLICATION NUMBER:  08/120,096
;      FILING DATE:  13-SEP-93
;      APPLICATION NUMBER:  08/111,519
;      FILING DATE:  25-AUG-93
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Buttmi, Jean A.
;      REGISTRATION NUMBER:  24,236
;      REFERENCE/DOCKET NUMBER:  A4710CIP3.SL3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)415-0100
;      TELEFAX:  (703)418-2768
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA viral
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Four Corners Hantavirus
;      INDIVIDUAL ISOLATE:  3H226
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:
;    PUBLICATION INFORMATION:
;      AUTHORS:  Hjelle, Brian
;      AUTHORS:  Jenison, Steven
;      AUTHORS:  Torrez-Martinez, No. 5837441ah
;      AUTHORS:  Yamada, Takashi
;      AUTHORS:  No. 5837441te, Kurt
;      AUTHORS:  Zumwalt, Ross
;      AUTHORS:  MacInnes, Kersti
;      AUTHORS:  Myers, Gerald
;      TITLE:  A No. 5837441el Hantavirus Associated with an Outbreak of Fatal Respir
;      TITLE:  Disease in the Southwestern United States: Evolutionary Relationships
;      TITLE:  Hantaviruses-Running Title: Hantavirus-associated ARDS
;      JOURNAL:  Journal of Virology
;      VOLUME:  68
;      PAGES:  in press
;      DATE:  1994
;      RELEVANT RESIDUES IN SEQ ID NO:  1:FROM 1 TO 24
US-08-210-762E-31
```

```
  Query Match            13.3%;  Score 12;  DB 2;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 gtttactgtaat 32
           ||||||||||||
Db       2 GTTTACTGTAAT 13


RESULT  27
US-08-256-799-26
; Sequence 26, Application US/08256799
; Patent No. 6222094
;  GENERAL INFORMATION:
;    APPLICANT:  HANSSON, Lennart
;    APPLICANT:  STROEMQVIST, Mats
;    APPLICANT:  BERGSTROEM, Sven
;    APPLICANT:  HERNELL, Olle
;    APPLICANT:  Toernell, Jan
;    TITLE OF INVENTION:  DNA ENCODING KAPPA-CASEIN, PROCESS FOR
;    TITLE OF INVENTION:  OBTAINING THE PROTEIN AND USE THEREOF
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/256,799
;      FILING DATE:  06-DEC-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 88/92
;      FILING DATE:  23-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  COOPER, Iver P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  HANSSON=1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-256-799-26


  Query Match            13.3%;  Score 12;  DB 4;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      35 ttatgttttcat 46
           ||||||||||||
Db      12 TTATGTTTTCAT 23


RESULT  28
US-08-462-437-26
; Sequence 26, Application US/08462437
; Patent No. 6232094
;  GENERAL INFORMATION:
;    APPLICANT:  HANSSON, Lennart
;    APPLICANT:  STROEMQVIST, Mats
;    APPLICANT:  BERGSTROEM, Sven
;    APPLICANT:  HERNELL, Olle
;    APPLICANT:  TOERNELL, Jan
;    TITLE OF INVENTION:  DNA ENCODING KAPPA-CASEIN, PROCESS
;    TITLE OF INVENTION:  FOR OBTAINING THE PROTEIN AND USE THEREOF
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,437
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 88/92
;      FILING DATE:  23-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  COOPER, Iver P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  HANSSON=1A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-462-437-26


  Query Match            13.3%;  Score 12;  DB 4;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      35 ttatgttttcat 46
           ||||||||||||
Db      12 TTATGTTTTCAT 23


RESULT  29
US-08-271-880A-140/c
; Sequence 140, Application US/08271880A
; Patent No. 5693535
;  GENERAL INFORMATION:
;    APPLICANT:  Kenneth G. Draper
;    APPLICANT:  Bharat Chowrira
;    APPLICANT:  James McSwiggen
;    APPLICANT:  Dan T. Stinchcomb
;    APPLICANT:  James D. Thompson
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR INHIBITING
;    TITLE OF INVENTION:  HUMAN IMMUNODEFICIENCY VIRUS
;    TITLE OF INVENTION:  REPLICATION
;    NUMBER OF SEQUENCES:  232
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
```

```
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/271,880A
;      FILING DATE:  July 7, 1994
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including appplication
;    PRIOR APPLICATION DATA:  described below:              two
;      APPLICATION NUMBER:  08/103,243
;      FILING DATE:  August 6, 1993
;      APPLICATION NUMBER:  07/882,886
;      FILING DATE:  May 14, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  206/116
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  140:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-271-880A-140


  Query Match             13.3%;  Score 12;  DB 1;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db      23 CAATGAAAGCAA 12


RESULT  30
US-08-910-408-140/c
; Sequence 140, Application US/08910408
; Patent No. 5972704
;  GENERAL INFORMATION:
;    APPLICANT:  Kenneth G. Draper
;    APPLICANT:  Bharat Chowrira
;    APPLICANT:  James McSwiggen
;    APPLICANT:  Dan T. Stinchcomb
;    APPLICANT:  James D. Thompson
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR INHIBITING
;    TITLE OF INVENTION:  HUMAN IMMUNODEFICIENCY VIRUS
;    TITLE OF INVENTION:  REPLICATION
;    NUMBER OF SEQUENCES:  232
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,408
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/271,880
;      FILING DATE:  July 7, 1994
;      APPLICATION NUMBER:  08/103,243
;      FILING DATE:  August 6, 1993
;      APPLICATION NUMBER:  07/882,886
;      FILING DATE:  May 14, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  206/116
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  140:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-910-408-140


  Query Match             13.3%;  Score 12;  DB 2;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db      23 CAATGAAAGCAA 12


RESULT  31
US-09-249-215-140/c
; Sequence 140, Application US/09249215
; Patent No. 6159692
;   GENERAL INFORMATION:
;        APPLICANT: Kenneth G. Draper
;                   Bharat Chowrira
;                   James McSwiggen
;                   Dan T. Stinchcomb
;                   James D. Thompson
;        TITLE OF INVENTION: METHOD AND REAGENT FOR INHIBITING
;                            HUMAN IMMUNODEFICIENCY VIRUS
;                            REPLICATION
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Lyon & Lyon
;             STREET: 633 West Fifth Street
;                     Suite 4700
;             CITY: Los Angeles
;             STATE: California
;             COUNTRY: U.S.A.
;             ZIP: 90071
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;                          storage
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: IBM P.C. DOS 5.0
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/249,215
;             FILING DATE: 12-Feb-1999
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/910,408
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/103,243
;             FILING DATE: August 6, 1993
;             APPLICATION NUMBER: 07/882,886
;             FILING DATE: May 14, 1992
;        ATTORNEY/AGENT INFORMATION:
```

```
;             NAME: Warburg, Richard
;             REGISTRATION NUMBER: 32,327
;             REFERENCE/DOCKET NUMBER: 206/116
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (213) 489-1600
;             TELEFAX: (213) 955-0440
;             TELEX: 67-3510
;    INFORMATION FOR SEQ ID NO: 140:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 25 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 140:
US-09-249-215-140


  Query Match           13.3%;  Score 12;  DB 3;  Length 25;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db      23 CAATGAAAGCAA 12


RESULT  32
US-08-835-728D-72
; Sequence 72, Application US/08835728D
; Patent No. 6017704
;  GENERAL INFORMATION:
;    APPLICANT:  Herman, James G.
;    APPLICANT:  Baylin, Stephen B.
;    TITLE OF INVENTION:  Methylation Specific Detection
;    NUMBER OF SEQUENCES:  216
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,728D
;      FILING DATE:  April 11, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/656,716
;      FILING DATE:  June 03, 1996,
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07265/125001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  72:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-835-728D-72


  Query Match           13.3%;  Score 12;  DB 3;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 tttttatgtttt 43
           ||||||||||||
Db       6 TTTTTATGTTTT 17


RESULT  33
US-08-835-728D-176/c
; Sequence 176, Application US/08835728D
; Patent No. 6017704
;  GENERAL INFORMATION:
;    APPLICANT:  Herman, James G.
;    APPLICANT:  Baylin, Stephen B.
;    TITLE OF INVENTION:  Methylation Specific Detection
;    NUMBER OF SEQUENCES:  216
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,728D
;      FILING DATE:  April 11, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/656,716
;      FILING DATE:  June 03, 1996,
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07265/125001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  176:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-835-728D-176


  Query Match           13.3%;  Score 12;  DB 3;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 tttttatgtttt 43
           ||||||||||||
Db      21 TTTTTATGTTTT 10


RESULT  34
US-09-490-558-72
; Sequence 72, Application US/09490558
; Patent No. 6265171
;  GENERAL INFORMATION:
;    APPLICANT:  Herman, James G.
;    APPLICANT:  Baylin, Stephen B.
;    TITLE OF INVENTION:  Methylation Specific Detection
;    NUMBER OF SEQUENCES:  216
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
```

```
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/490,558
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/835,728
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07265/125001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  72:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-490-558-72


  Query Match             13.3%;  Score 12;  DB 4;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 tttttatgtttt 43
           ||||||||||||
Db       6 TTTTTATGTTTT 17


RESULT  35
US-09-490-558-176/c
; Sequence 176, Application US/09490558
; Patent No. 6265171
;  GENERAL INFORMATION:
;    APPLICANT:  Herman, James G.
;    APPLICANT:  Baylin, Stephen B.
;    TITLE OF INVENTION:  Methylation Specific Detection
;    NUMBER OF SEQUENCES:  216
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/490,558
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/835,728
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07265/125001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  176:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-490-558-176


  Query Match             13.3%;  Score 12;  DB 4;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 tttttatgtttt 43
           ||||||||||||
Db      21 TTTTTATGTTTT 10


RESULT  36
US-08-668-128B-6/c
; Sequence 6, Application US/08668128B
; Patent No. 5840568
;  GENERAL INFORMATION:
;    APPLICANT:  Pfreundschuh, Michael
;    TITLE OF INVENTION:  Hodgkin's Disease Associated Molecules And
;    TITLE OF INVENTION:  Uses Thereof
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,128B
;      FILING DATE:  21-JUNE-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/644,116
;      FILING DATE:  10-MAY-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/580,980
;      FILING DATE:  03-JANUARY-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/479,328
;      FILING DATE:  07-JUNE-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5840568man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5441
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-668-128B-6


  Query Match             13.3%;  Score 12;  DB 2;  Length 27;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      77 tgaaagcaagtg 88
           ||||||||||||
Db      26 TGAAAGCAAGTG 15


RESULT  37
US-08-905-445-6/c
; Sequence 6, Application US/08905445
; Patent No. 5864015
;  GENERAL INFORMATION:
;    APPLICANT:  Pfreundschuh, Michael
;    TITLE OF INVENTION:  Hodgkin's Disease Associated Molecules And
;    TITLE OF INVENTION:  Uses Thereof
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,445
;      FILING DATE:  04-AUG-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/668,128
;      FILING DATE:  21-JUNE-1996
;      APPLICATION NUMBER:  08/644,116
;      FILING DATE:  10-MAY-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/580,980
;      FILING DATE:  03-JANUARY-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/479,328
;      FILING DATE:  07-JUNE-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5864015man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5441
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-905-445-6


  Query Match             13.3%;  Score 12;  DB 2;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      77 tgaaagcaagtg 88
           ||||||||||||
Db      26 TGAAAGCAAGTG 15


RESULT  38
US-08-644-116A-6/c
; Sequence 6, Application US/08644116A
; Patent No. 6140464
;  GENERAL INFORMATION:
;    APPLICANT:  Pfreundschuh, Michael; Rammensee, Hans-Georg
;    TITLE OF INVENTION:  Method For Identifying Or Isolating A Molecule
;    TITLE OF INVENTION:  And Molecules Identified Thereby
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,116A
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/580,980
;      FILING DATE:  03-JANUARY-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/479,328
;      FILING DATE:  07-JUNE-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6140464man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5410.2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-644-116A-6


  Query Match             13.3%;  Score 12;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      77 tgaaagcaagtg 88
           ||||||||||||
Db      26 TGAAAGCAAGTG 15


RESULT  39
US-08-771-602D-20
; Sequence 20, Application US/08771602D
; Patent No. 5976795
;  GENERAL INFORMATION:
;    APPLICANT:  Voytas, Daniel F.
;    APPLICANT:  Zou, Sige
;    TITLE OF INVENTION:  Retrotransposon and Methods
;    NUMBER OF SEQUENCES:  51
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/771,602D
```

```
;      FILING DATE:  20-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/010,869
;      FILING DATE:  31-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  8-96
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Oligonucleotide."
;    HYPOTHETICAL:  NO
US-08-771-602D-20


  Query Match             13.3%;  Score 12;  DB 2;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 atttttatgttt 42
           ||||||||||||
Db       7 ATTTTTATGTTT 18


RESULT  40
US-08-771-602D-21/c
; Sequence 21, Application US/08771602D
; Patent No. 5976795
;  GENERAL INFORMATION:
;    APPLICANT:  Voytas, Daniel F.
;    APPLICANT:  Zou, Sige
;    TITLE OF INVENTION:  Retrotransposon and Methods
;    NUMBER OF SEQUENCES:  51
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/771,602D
;      FILING DATE:  20-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/010,869
;      FILING DATE:  31-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  8-96
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Oligonucleotide."
;    HYPOTHETICAL:  NO
US-08-771-602D-21


  Query Match             13.3%;  Score 12;  DB 2;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 atttttatgttt 42
           ||||||||||||
Db      22 ATTTTTATGTTT 11


RESULT  41
US-08-619-724-3/c
; Sequence 3, Application US/08619724
; Patent No. 5827653
;  GENERAL INFORMATION:
;    APPLICANT:  SAMMES, Peter George
;    APPLICANT:  GARMAN, Andrew John
;    TITLE OF INVENTION:  NUCLEIC ACID DETECTION WITH ENERGY TRANSFER
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY MADISON & SUTRO, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,724
;      FILING DATE:  20-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB94/02068
;      FILING DATE:  23-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9412106.8
;      FILING DATE:  16-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9319826.5
;      FILING DATE:  23-SEP-1993
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-619-724-3


  Query Match             13.3%;  Score 12;  DB 1;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaattttta 37
           ||||||||||||
Db      29 CTGTAATTTTTA 18


RESULT  42
US-08-629-001A-25/c
; Sequence 25, Application US/08629001A
; Patent No. 5858661
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Shiloh, Yosef
;    TITLE OF INVENTION:  ATAXIA-TELANGIECTASIA GENE AND ITS
;    TITLE OF INVENTION:  GENOMIC ORGANIZATION
;    NUMBER OF SEQUENCES:  139
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 5858661thwestern Hwy.
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,001A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohn, Kenneth I.
;      REGISTRATION NUMBER:  30,955
;      REFERENCE/DOCKET NUMBER:  2290.00032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 539-5050
;      TELEFAX:  (810) 539-5055
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-629-001A-25


  Query Match           13.3%;  Score 12;  DB 2;  Length 30;
  Best Local Similarity  100.0%;  Pred. No. 2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatt 15
          ||||||||||||
Db     19 AAAAAAATTATT 8


RESULT  43
US-08-642-274D-104/c
; Sequence 104, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
;  APPLICANT:  Shiloh, Yosef
;  TITLE OF INVENTION:  MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
;  TITLE OF INVENTION:  SCREEN FOR A PARTIAL A-T PHENOTYPE
;  FILE REFERENCE:  229000033
;  CURRENT APPLICATION NUMBER: US/08/642,274D
;  CURRENT FILING DATE:  1996-05-03
;  NUMBER OF SEQ ID NOS:  220
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 104
;   LENGTH: 30
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:intronic
;   OTHER INFORMATION: sequence
US-08-642-274D-104


  Query Match           13.3%;  Score 12;  DB 4;  Length 30;
  Best Local Similarity  100.0%;  Pred. No. 2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatt 15
          ||||||||||||
Db     19 AAAAAAATTATT 8


RESULT  44
US-08-935-312-14/c
; Sequence 14, Application US/08935312
; Patent No. 6207455
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, Lung-Ji
;    TITLE OF INVENTION:  LENTIVIRAL VECTORS
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;      STREET:  624 Ninth Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/935,312
;      FILING DATE:  22-SEP-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  COOPER, Iver P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  CHANG=112
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = ""DNA""
US-08-935-312-14


  Query Match           13.3%;  Score 12;  DB 4;  Length 30;
  Best Local Similarity  100.0%;  Pred. No. 2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     74 caatgaaagcaa 85
          ||||||||||||
Db     28 CAATGAAAGCAA 17


RESULT  45
US-08-341-456A-14/c
; Sequence 14, Application US/08341456A
; Patent No. 5767074
;  GENERAL INFORMATION:
;    APPLICANT:  Besmer, Peter
;    APPLICANT:  No. 5767074ka, Karl
;    APPLICANT:  Buck, Jochen
;    APPLICANT:  Moore, Malcolm A.S.
;    TITLE OF INVENTION:  COMPOSITIONS OF SOLUBLE C-KIT LIGAND AND
;    TITLE OF INVENTION:  HEMATOPOIETIC FACTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
```

```
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/341,456A
;      FILING DATE:  17-NOV-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  37454-C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-341-456A-14


  Query Match             13.3%;  Score 12;  DB 1;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 ttcatgttttct 53
           ||||||||||||
Db      21 TTCATGTTTTCT 10



Search completed: January 24, 2002, 03:24:26
Job time: 3713 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:57:29 ; Search time 2099.46 Seconds
(without alignments)
460.652 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 80718

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	14	15.6	50	13	TA154F10Q	AL473287 T. brucei
c	2	13	14.4	25	13	AZ829725	AZ829725 2M0107I24
c	3	13	14.4	26	13	TA123B12Q	AL463522 T. brucei
	4	13	14.4	28	13	AZ452653	AZ452653 1M0252E07
	5	13	14.4	36	13	AZ314238	AZ314238 1M0030N24
	6	13	14.4	49	10	AA500776	AA500776 vg01b11.r
	7	13	14.4	50	10	AU102750	AU102750 AU102750
c	8	12	13.3	23	13	AZ579583	AZ579583 1M0367N03
c	9	12	13.3	24	13	AZ996858	AZ996858 2M0283L08
c	10	12	13.3	25	13	AZ345616	AZ345616 1M0080E20
	11	12	13.3	27	10	AW246455	AW246455 2821693.3
	12	12	13.3	30	13	AZ610538	AZ610538 1M0435E23
	13	12	13.3	33	13	AZ507698	AZ507698 1M0349O21
	14	12	13.3	38	13	AZ487251	AZ487251 1M0316A18
c	15	12	13.3	41	13	AZ662545	AZ662545 1M0541P07
c	16	12	13.3	42	11	T17635	T17635 mps v360 Th
	17	12	13.3	45	13	AZ514394	AZ514394 1M0361J02
	18	12	13.3	46	10	AI805932	AI805932 te52g03.x
c	19	12	13.3	48	13	AZ832106	AZ832106 2M0112I13
c	20	12	13.3	50	11	D19972	D19972 HUMGS00937
	21	11	12.2	20	13	AZ782616	AZ782616 2M0023F17
	22	11	12.2	21	13	AZ768088	AZ768088 1M0567G21
c	23	11	12.2	22	10	AA991150	AA991150 os40a07.s
	24	11	12.2	22	13	AZ766483	AZ766483 1M0564A10
	25	11	12.2	23	13	AZ387817	AZ387817 1M0147B24
	26	11	12.2	24	13	TA242F03P	AL482984 T. brucei
	27	11	12.2	25	13	AZ462654	AZ462654 1M0269P09
c	28	11	12.2	28	13	AZ809415	AZ809415 2M0073I20
	29	11	12.2	29	11	D18726	D18726 MUSGS01788
	30	11	12.2	29	13	AZ580321	AZ580321 1M0368G02
	31	11	12.2	31	13	AZ579378	AZ579378 1M0363N13
c	32	11	12.2	32	13	AZ331642	AZ331642 1M0059P11
	33	11	12.2	32	13	AZ591923	AZ591923 1M0402F21
c	34	11	12.2	33	11	U44209	U44209 ENU44209 As
	35	11	12.2	33	13	AZ423204	AZ423204 1M0202O10
	36	11	12.2	33	13	AZ494328	AZ494328 1M0329F05
	37	11	12.2	35	10	AU014463	AU014463 AU014463
c	38	11	12.2	35	11	D19995	D19995 HUMGS00963
c	39	11	12.2	36	10	AU111149	AU111149 AU111149
c	40	11	12.2	36	13	AZ381596	AZ381596 1M0138C16
c	41	11	12.2	36	13	AZ807406	AZ807406 2M0070D15
	42	11	12.2	37	10	AA647854	AA647854 vq80e05.s
	43	11	12.2	37	10	AI200438	AI200438 qf93b01.x
	44	11	12.2	37	13	AZ806836	AZ806836 2M0069I08
c	45	11	12.2	38	11	D21038	D21038 HUMGS02021

ALIGNMENTS

RESULT 1
TA154F10Q
LOCUS TA154F10Q 50 bp DNA GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 154f10, reverse sequence, genomic survey sequence.
ACCESSION AL473287
VERSION AL473287.1 GI:11838560
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 50)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh1@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (
4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES Location/Qualifiers
source 1. .50

```
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="154f10"
BASE COUNT       20 a      7 c      9 g     14 t
ORIGIN


  Query Match             15.6%;  Score 14;  DB 13;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgaaaaaaattat 14
          ||||||||||||||
Db     10 ATGAAAAAAATTAT 23


RESULT   2
AZ829725/c
LOCUS       AZ829725       25 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0107I24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0107I24 F, DNA sequence.
ACCESSION   AZ829725
VERSION     AZ829725.1  GI:12999549
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0107  row: I  column: 24
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 25.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0107I24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      3 c      4 g      4 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 13;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 9.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     34 tttatgttttcat 46
          |||||||||||||
Db     24 TTTATGTTTTCAT 12


RESULT   3
TA123B12Q/c
LOCUS       TA123B12Q      26 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 123b12, reverse sequence,
            genomic survey sequence.
ACCESSION   AL463522
VERSION     AL463522.1  GI:11834032
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="123b12"
BASE COUNT       12 a      0 c      0 g     14 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 13;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 9.5e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattattt 16
          |||||||||||||
Db     20 AAAAAAATTATTT 8


RESULT   4
AZ452653
LOCUS       AZ452653       28 bp    DNA             GSS       04-OCT-2000
```

```
DEFINITION  1M0252E07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0252E07 R, DNA sequence.
ACCESSION   AZ452653
VERSION     AZ452653.1  GI:10609676
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0252  row: E  column: 07
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0252E07"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      2 c      2 g     10 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 13;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 aaaaaaattattt 16
           |||||||||||||
Db      16 AAAAAAATTATTT 28


RESULT   5
AZ314238
```

```
LOCUS       AZ314238       36 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0030N24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0030N24 R, DNA sequence.
ACCESSION   AZ314238
VERSION     AZ314238.1  GI:10359929
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0030  row: N  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 36.
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0030N24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        8 a      3 c      5 g     20 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 13;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      45 atgttttcttatt 57
           |||||||||||||
Db       2 ATGTTTTCTTATT 14


RESULT   6
```

```
AA500776
LOCUS       AA500776       49 bp    mRNA            EST       01-JUL-1997
DEFINITION  vg01b11.r1 Soares mouse NbMH Mus musculus cDNA clone IMAGE:860061
            5' similar to TR:G1333638 G1333638 PARAOXONASE 2. ;, mRNA sequence.
ACCESSION   AA500776
VERSION     AA500776.1  GI:2235743
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:504149
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Mus musculus".
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:860061"
                     /clone_lib="Soares mouse NbMH"
                     /sex="male"
                     /tissue_type="heart"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAAGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       17 a      8 c     10 g     14 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 10;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 7.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      30 aatttttatgttt 42
           |||||||||||||
Db      25 AATTTTTATGTTT 37


RESULT   7
AU102750
LOCUS       AU102750       50 bp    mRNA            EST       05-APR-2001
DEFINITION  AU102750 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            CAS10730, mRNA sequence.
ACCESSION   AU102750
VERSION     AU102750.1  GI:13552271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,K., Suyama,A. and Sugano,S.
  TITLE     Fine Structural analysis of transcription start sites of human
            mRNAs using full-length enriched and 5'-end enriched cDNA libraries
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CAS10730"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       19 a      13 c      8 g      10 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 10;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 7.6e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      73 ccaatgaaagcaa 85
           |||||||||||||
Db       8 CCAATGAAAGCAA 20


RESULT   8
AZ579583/c
LOCUS       AZ579583       23 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0367N03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0367N03 F, DNA sequence.
ACCESSION   AZ579583
VERSION     AZ579583.1  GI:11694012
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0367  row: N  column: 03
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 23.
```

```
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0367N03"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        9 a      2 c      0 g     12 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 aaaaaaattatt 15
           ||||||||||||
Db      15 AAAAAAATTATT 4


RESULT   9
AZ996858/c
LOCUS       AZ996858       24 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0283L08F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0283L08 F, DNA sequence.
ACCESSION   AZ996858
VERSION     AZ996858.1  GI:13868085
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0283  row: L  column: 08
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 24.
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0283L08"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       12 a      4 c      3 g      5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      33 ttttatgttttc 44
           ||||||||||||
Db      21 TTTTATGTTTTC 10


RESULT  10
AZ345616/c
LOCUS       AZ345616       25 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0080E20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0080E20 F, DNA sequence.
ACCESSION   AZ345616
VERSION     AZ345616.1  GI:10424853
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0080  row: E  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
```

```
          Class: plasmid ends
          High quality sequence stop: 25.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0080E20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       16 a      1 c      0 g      8 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      29 taatttttatgt 40
           ||||||||||||
Db      12 TAATTTTTATGT 1


RESULT  11
AW246455
LOCUS       AW246455       27 bp    mRNA            EST       07-JAN-2000
DEFINITION  2821693.3prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821693 3',
            mRNA sequence.
ACCESSION   AW246455
VERSION     AW246455.1  GI:6589448
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: 2821693.5prime
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling
            Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
            Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
            project Clone distribution: MGC clone distribution information can
            be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality
            Scores: PHRED from University of Washington Genome Center. Vector
            Trimming: cross_match from University of Washington Genome Center
            PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley
            Drosophila Genome Project. University of Washington Genome Center:
            http://www.genome.washington.edu Low Quality Sequence: 10
            contiguous PHRED high quality bases following vector sequence. Very
            Low Quality Sequence: Trace file contained 27 contiguous distinct
            peaks following vector sequence. Polyadenylation: Based upon the
            presence of a XhoI site followed by a run of 14 or more T residues
            at the beginning of the sequence, this cDNA insert was
            polyadenylated.
            Plate: LLCM7  row: H  column: 14
            High quality sequence stop: 10.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2821693"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       2 a      1 c      2 g     22 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 10;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      40 ttttcatgtttt 51
           ||||||||||||
Db       7 TTTTCATGTTTT 18


RESULT  12
AZ610538
LOCUS       AZ610538       30 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0435E23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0435E23 R, DNA sequence.
ACCESSION   AZ610538
VERSION     AZ610538.1  GI:11732728
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0435  row: E  column: 23
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 30.
FEATURES             Location/Qualifiers
     source          1. .30
```

```
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0435E23"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       10 a      2 c      4 g     14 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      27 tgtaatttttat 38
           ||||||||||||
Db       4 TGTAATTTTTAT 15


RESULT  13
AZ507698
LOCUS       AZ507698       33 bp    DNA             GSS       05-OCT-2000
DEFINITION  1M0349O21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0349O21 F, DNA sequence.
ACCESSION   AZ507698
VERSION     AZ507698.1  GI:10689014
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0349  row: O  column: 21
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 33.
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0349O21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       4 a      3 c      6 g     20 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      40 ttttcatgtttt 51
           ||||||||||||
Db      16 TTTTCATGTTTT 27


RESULT  14
AZ487251
LOCUS       AZ487251       38 bp    DNA             GSS       05-OCT-2000
DEFINITION  1M0316A18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0316A18 R, DNA sequence.
ACCESSION   AZ487251
VERSION     AZ487251.1  GI:10654814
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0316  row: A  column: 18
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 38.
```

```
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0316A18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      14 c       5 g       5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 tgaaaaaaatta 13
           ||||||||||||
Db      26 TGAAAAAAATTA 37


RESULT  15
AZ662545/c
LOCUS       AZ662545         41 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0541P07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0541P07 R, DNA sequence.
ACCESSION   AZ662545
VERSION     AZ662545.1  GI:11799691
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0541  row: P  column: 07
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 41.
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0541P07"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       16 a      9 c       5 g      11 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      52 cttattgttgga 63
           ||||||||||||
Db      35 CTTATTGTTGGA 24


RESULT  16
T17635/c
LOCUS       T17635           42 bp    mRNA            EST       06-JUN-1994
DEFINITION  mps v360 The blue guys library Saccharomyces cerevisiae cDNA
            sequence upstream of LacZ fusion similar to GAC1, X63941, mRNA
            sequence.
ACCESSION   T17635
VERSION     T17635.1  GI:459560
KEYWORDS    EST.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Burns,N., Grimwade,B., Ross-Macdonald,P.B., Choi,E.-Y., Finberg,K.,
            Roeder,G.S. and Snyder,M.
  TITLE     Large-scale analysis of gene expression, protein localization and
            gene disruption in Saccharomyces cerevisiae
  JOURNAL   Genes Dev. 8, 1087-1105 (1994)
  MEDLINE   95011603
COMMENT     Contact: Snyder M
            Department of Biology
            Yale University
            New Haven CT 06520-8103
            Tel: 2034326139
            Fax: 2034326161
            Email: snymicp@yalevm.ycc.yale.edu
            LacZ fusion; Vegetative expression; Beta-gal fusion  localization
            pattern:
            50 cytoplasmic spots; Disruption  phenotype: none detected; Fusion:
            codon 407 of GAC1 gene.   Sequence below near or adjacent to lacZ.
```

```
          Seq primer: LacZ sequences in transposon.
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /clone_lib="The blue guys library"
                     /lab_host="E.coli"
                     /note="Vector: pRECmTn; A yeast genomic DNA library was
                     prepared in the vector pHSS6, and subjected to transposon
                     mutagenesis with mTn3. This mini-transposon carries lacZ
                     sequences that lack an initiation codon; expression of
                     lacZ is only provided by in frame fusion to yeast coding
                     sequence. The yeast genomic DNA carrying the transposon
                     was excised from pHSS6 and transplaced back onto the yeast
                     chromosome. Yeast colonies expressing lacZ were screened
                     for in a color assay. A plasmid containing the genomic
                     DNA/lacZ fusion junction was recovered from each
                     individual yeast colony that expressed lacZ activity.
                     These recovered plasmids comprise 'The blue guys library'.
                     The fusion junction was then sequenced to identify the
                     expressed ORF upstream of the fusion. "
BASE COUNT        7 a     12 c     11 g     12 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 11;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 aatgaaagcaag 86
           ||||||||||||
Db      22 AATGAAAGCAAG 11


RESULT  17
AZ514394
LOCUS       AZ514394      45 bp    DNA              GSS       05-OCT-2000
DEFINITION  1M0361J02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0361J02 F, DNA sequence.
ACCESSION   AZ514394
VERSION     AZ514394.1  GI:10695710
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0361  row: J  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 45.
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0361J02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       24 a      2 c      9 g     10 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 13;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 aaaaaaattatt 15
            ||||||||||||
Db       3 AAAAAAATTATT 14


RESULT  18
AI805932
LOCUS       AI805932      46 bp    mRNA             EST       13-DEC-1999
DEFINITION  te52g03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2090356 3' similar to TR:Q99766 Q99766 HYPOTHETICAL 15.7 KD
            PROTEIN. ;, mRNA sequence.
ACCESSION   AI805932
VERSION     AI805932.1  GI:5392498
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 606   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2090356"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
```

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 13 a 9 c 5 g 19 t
ORIGIN

Query Match 13.3%; Score 12; DB 10; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.3e+05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 actgtaattttt 36
||||||||||||
Db 34 ACTGTAATTTTT 45

RESULT 19
AZ832106/c
LOCUS AZ832106 48 bp DNA GSS 20-FEB-2001
DEFINITION 2M0112I13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0112I13 F, DNA sequence.
ACCESSION AZ832106
VERSION AZ832106.1 GI:13002014
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 48)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0112 row: I column: 13
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 48.
FEATURES Location/Qualifiers
source 1. .48
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0112I13"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 14 a 11 c 3 g 20 t
ORIGIN

Query Match 13.3%; Score 12; DB 13; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aaaaaaattatt 15
||||||||||||
Db 48 AAAAAAATTATT 37

RESULT 20
D19972/c
LOCUS D19972 50 bp mRNA EST 30-JUL-1996
DEFINITION HUMGS00937 Human promyelocyte Homo sapiens cDNA clone mm06d08 3', mRNA sequence.
ACCESSION D19972
VERSION D19972.1 GI:500869
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 50)
AUTHORS Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y., Yoshinari,H., Arimoto,J. and Matsubara,K.
TITLE Gene expression of human promyelocytic cell line HL60 before and after induction of differentiation. A new application of 3'directed cDNA sequencing
JOURNAL Unpublished (1993)
COMMENT Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y., Yoshinari,H., Arimoto,J. and Matsubara,K.
Institute for Molecular and Cellular Biology
Osaka University
3-1 Yamada-oka,Suita,Osaka 565,Japan.
FEATURES Location/Qualifiers
source 1. .50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="mm06d08"
/clone_lib="Human promyelocyte"
/note="Female, adult, cell_line = HL60, cell_type = promyelocyte. "

BASE COUNT 20 a 5 c 11 g 14 t
ORIGIN

Query Match 13.3%; Score 12; DB 11; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 ttcaaagtttac 26
||||||||||||
Db 20 TTCAAAGTTTAC 9

RESULT 21
AZ782616
LOCUS AZ782616 20 bp DNA GSS 16-FEB-2001
DEFINITION 2M0023F17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0023F17 R, DNA sequence.
ACCESSION AZ782616
VERSION AZ782616.1 GI:12916517

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 20)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0023 row: F column: 17
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 20.

FEATURES Location/Qualifiers
source 1. .20
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0023F17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 6 a 4 c 1 g 9 t
ORIGIN

Query Match 12.2%; Score 11; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.8e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33 ttttatgtttt 43
||||||||||||
Db 9 TTTTATGTTTT 19

RESULT 22
AZ768088
LOCUS AZ768088 21 bp DNA GSS 16-FEB-2001
DEFINITION 1M0567G21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0567G21 R, DNA sequence.
ACCESSION AZ768088

VERSION AZ768088.1 GI:12886839
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 21)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0567 row: G column: 21
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 21.

FEATURES Location/Qualifiers
source 1. .21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0567G21"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 3 c 1 g 12 t
ORIGIN

Query Match 12.2%; Score 11; DB 13; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 ttttcatgttt 50
||||||||||||
Db 11 TTTTCATGTTT 21

RESULT 23
AA991150/c
LOCUS AA991150 22 bp mRNA EST 03-JUN-1998
DEFINITION os40a07.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1607796 3' similar to TR:Q35990 Q35990 HYPOTHETICAL 8.9 KD PROTEIN. ;contains

```
          element L1 repetitive element ;, mRNA sequence.
ACCESSION   AA991150
VERSION     AA991150.1  GI:3177639
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1607796"
                     /clone_lib="NCI_CGAP_Br2"
                     /sex="female, pooled"
                     /tissue_type="breast"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from pooled bulk
                     breast tumor tissue, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  This library is the normalized version of
                     NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                     and M. Fatima Bonaldo. "
BASE COUNT       16 a      3 c      2 g      1 t
ORIGIN


  Query Match           12.2%;  Score 11;  DB 10;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      46 tgttttcttat 56
           ||||||||||
Db      22 TGTTTTCTTAT 12


RESULT  24
AZ766483
LOCUS       AZ766483       22 bp    DNA             GSS       16-FEB-2001
DEFINITION  1M0564A10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0564A10 F, DNA sequence.
ACCESSION   AZ766483
VERSION     AZ766483.1  GI:12883604
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0564  row: A  column: 10
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 22.
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0564A10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        4 a      2 c      4 g     12 t
ORIGIN


  Query Match           12.2%;  Score 11;  DB 13;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      46 tgttttcttat 56
           ||||||||||
Db       4 TGTTTTCTTAT 14


RESULT  25
AZ387817
LOCUS       AZ387817       23 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0147B24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0147B24 R, DNA sequence.
ACCESSION   AZ387817
VERSION     AZ387817.1  GI:10501525
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
```

```
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0147  row: B  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 23.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0147B24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        5 a      4 c      3 g     11 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 8.4e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      28 gtaatttttat 38
           ||||||||||
Db       6 GTAATTTTTAT 16


RESULT  26
TA242F03P
LOCUS       TA242F03P      24 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 242f03, forward sequence,
            genomic survey sequence.
ACCESSION   AL482984
VERSION     AL482984.1  GI:11848725
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="242f03"
BASE COUNT        1 a      3 c     10 g     10 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 8.2e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      52 cttattgttgg 62
           ||||||||||
Db       8 CTTATTGTTGG 18


RESULT  27
AZ462654
LOCUS       AZ462654       25 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0269P09R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0269P09 R, DNA sequence.
ACCESSION   AZ462654
VERSION     AZ462654.1  GI:10620695
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0269  row: P  column: 09
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 25.
FEATURES             Location/Qualifiers
```

```
    source          1. .25
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGC1M0269P09"
                    /clone_lib="Mouse 10kb plasmid UUGC1M library"
                    /sex="Male"
                    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                    /note="Vector: PWD42nv; Purified genomic DNA from M.
                    musculus C57BL/6J (male) was obtained  from the Jackson
                    Laboratory Mouse DNA Resource
                    (http://www.jax.org/resources/documents/dnares/). The DNA
                    was hydrodynamically sheared by repeated passage through a
                    0.005 inch orifice at constant velocity. The sheared DNA
                    was blunt end-repaired with T4 DNA polymerase and T4
                    polynucleotide kinase. Adaptor oligonucleotides were
                    ligated to the blunt ends in high  molar excess. The
                    adaptored DNA was purified and size-selected for a 9.5 to
                    10.5 kb range using preparative agarose gel
                    electrophoresis. Vector DNA was prepared from a derivative
                    of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                    inducible derivative of plasmid R1. The vector was ligated
                    with adaptors complementary to the insert adaptors and
                    purified. The sheared, adaptored mouse DNA was annealed to
                    adaptored vector DNA, and transformed into
                    chemically-competent E. coli XL10-Gold (Stratagene) cells
                    and selected for ampicillin resistance."
BASE COUNT       10 a      2 c      0 g     13 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 8.1e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aaaaattattt 16
          ||||||||||||
Db      4 AAAAATTATTT 14


RESULT  28
AZ809415/c
LOCUS       AZ809415         28 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0073I20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0073I20 F, DNA sequence.
ACCESSION   AZ809415
VERSION     AZ809415.1  GI:12975693
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0073  row: I  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES            Location/Qualifiers
     source         1. .28
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGC2M0073I20"
                    /clone_lib="Mouse 10kb plasmid UUGC1M library"
                    /sex="Male"
                    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                    /note="Vector: PWD42nv; Purified genomic DNA from M.
                    musculus C57BL/6J (male) was obtained  from the Jackson
                    Laboratory Mouse DNA Resource
                    (http://www.jax.org/resources/documents/dnares/). The DNA
                    was hydrodynamically sheared by repeated passage through a
                    0.005 inch orifice at constant velocity. The sheared DNA
                    was blunt end-repaired with T4 DNA polymerase and T4
                    polynucleotide kinase. Adaptor oligonucleotides were
                    ligated to the blunt ends in high  molar excess. The
                    adaptored DNA was purified and size-selected for a 9.5 to
                    10.5 kb range using preparative agarose gel
                    electrophoresis. Vector DNA was prepared from a derivative
                    of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                    inducible derivative of plasmid R1. The vector was ligated
                    with adaptors complementary to the insert adaptors and
                    purified. The sheared, adaptored mouse DNA was annealed to
                    adaptored vector DNA, and transformed into
                    chemically-competent E. coli XL10-Gold (Stratagene) cells
                    and selected for ampicillin resistance."
BASE COUNT       7 a      5 c      5 g     11 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagca 84
           ||||||||||||
Db      14 CAATGAAAGCA 4


RESULT  29
D18726
LOCUS       D18726         29 bp    mRNA            EST       12-DEC-1995
DEFINITION  MUSGS01788 Mouse 3'-directed Mus musculus domesticus cDNA clone
            md0169 3', mRNA sequence.
ACCESSION   D18726
VERSION     D18726.1  GI:1100695
KEYWORDS    EST.
SOURCE      western European house mouse.
  ORGANISM  Mus musculus domesticus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara,K.
  TITLE     Analysis of gene expression in mouse embryogenesis by 3'-directed
            cDNA sequencing
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara
            ,K.
            Institute for Cellular and Molecular Biology
            Osaka University
            3-1 Yamada-oka, Suita, Osaka 565, Japan.
FEATURES            Location/Qualifiers
     source         1. .29
                    /organism="Mus musculus domesticus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10092"
                    /clone="md0169"
                    /clone_lib="Mouse 3'-directed"
                    /tissue_type="decidual tissue (day 6.5-8.5 of gestation)"
BASE COUNT      15 a      3 c      2 g      9 t
ORIGIN
```

```
  Query Match             12.2%;  Score 11;  DB 11;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 7.7e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      36 tatgttttcat 46
           ||||||||||
Db      15 TATGTTTTCAT 25


RESULT  30
AZ580321
LOCUS       AZ580321       29 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0368G02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0368G02 R, DNA sequence.
ACCESSION   AZ580321
VERSION     AZ580321.1  GI:11694750
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0368  row: G  column: 02
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 29.
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0368G02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       10 a     11 c      3 g      5 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 7.7e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      76 atgaaagcaag 86
           ||||||||||
Db      16 ATGAAAGCAAG 26


RESULT  31
AZ579378
LOCUS       AZ579378       31 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0363N13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0363N13 R, DNA sequence.
ACCESSION   AZ579378
VERSION     AZ579378.1  GI:11693807
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 31)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0363  row: N  column: 13
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 31.
FEATURES             Location/Qualifiers
     source          1. .31
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0363N13"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
```

```
BASE COUNT       12 a      1 c      8 g     10 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 7.5e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattat 14
          ||||||||||
Db      3 AAAAAAATTAT 13


RESULT  32
AZ331642/c
LOCUS       AZ331642       32 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0059P11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0059P11 R, DNA sequence.
ACCESSION   AZ331642
VERSION     AZ331642.1  GI:10394528
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0059  row: P  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 32.
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0059P11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      7 c      1 g     10 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 7.5e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     13 atttcaaagtt 23
          ||||||||||
Db     20 ATTTCAAAGTT 10


RESULT  33
AZ591923
LOCUS       AZ591923       32 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0402F21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0402F21 F, DNA sequence.
ACCESSION   AZ591923
VERSION     AZ591923.1  GI:11714113
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0402  row: F  column: 21
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 32.
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0402F21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
```

chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 8 a 5 c 9 g 10 t
ORIGIN

Query Match 12.2%; Score 11; DB 13; Length 32;
Best Local Similarity 100.0%; Pred. No. 7.5e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 aaagtttactg 28
||||||||||||
Db 4 AAAGTTTACTG 14

RESULT 34
U44209/c
LOCUS U44209 33 bp mRNA EST 03-APR-1996
DEFINITION ENU44209 Aspergillus nidulans cleistothecium Emericella nidulans
cDNA clone SE0393, mRNA sequence.
ACCESSION U44209
VERSION U44209.1 GI:1244872
KEYWORDS EST.
SOURCE Emericella nidulans.
ORGANISM Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
REFERENCE 1 (bases 1 to 33)
AUTHORS Lee,D., Lee,S., Hwang,H., Kim,J. and Chae,K.
TITLE Quantitative analysis of gene expression in sexual structures of
Aspergillus nidulans by sequencing of 3'-directed cDNA clones
JOURNAL FEMS Microbiol. Lett. 138 (1), 71-76 (1996)
MEDLINE 96236220
COMMENT Contact: Keon-Sang Chae
Chonbuk National University
Chonju, 561-756, S. Korea
Tel: +82-652-70-3340
Fax: +82-652-70-3345
Email: chaeks@chonbuknms.chonbuk.ac.kr.
FEATURES Location/Qualifiers
source 1. .33
/organism="Emericella nidulans"
/strain="FGSC4"
/db_xref="taxon:162425"
/clone="SE0393"
/clone_lib="Aspergillus nidulans cleistothecium"
/tissue_type="cleistothecium"
/cell_type="Hull cell"
/dev_stage="sexual"
/note="3'-directed cDNA clones; single-pass sequencing"
BASE COUNT 18 a 3 c 2 g 10 t
ORIGIN

Query Match 12.2%; Score 11; DB 11; Length 33;
Best Local Similarity 100.0%; Pred. No. 7.4e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 tttttatgttt 42
||||||||||||
Db 31 TTTTTATGTTT 21

RESULT 35
AZ423204
LOCUS AZ423204 33 bp DNA GSS 03-OCT-2000
DEFINITION 1M0202010F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0202010 F, DNA sequence.
ACCESSION AZ423204
VERSION AZ423204.1 GI:10547217
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 33)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0202 row: O column: 10
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 33.
FEATURES Location/Qualifiers
source 1. .33
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0202O10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 15 a 3 c 7 g 8 t
ORIGIN

Query Match 12.2%; Score 11; DB 13; Length 33;
Best Local Similarity 100.0%; Pred. No. 7.4e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aaaaaaattat 14
||||||||||||
Db 21 AAAAAAATTAT 31

RESULT 36
AZ494328
LOCUS AZ494328 33 bp DNA GSS 05-OCT-2000
DEFINITION 1M0329F05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0329F05 R, DNA sequence.
ACCESSION AZ494328
VERSION AZ494328.1 GI:10668799
KEYWORDS GSS.
SOURCE house mouse.

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0329  row: F  column: 05
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 33.
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0329F05"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        8 a      1 c      7 g     17 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 7.4e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 tttcttattgt 59
           ||||||||||||
Db      23 TTTCTTATTGT 33


RESULT  37
AU014463
LOCUS       AU014463       35 bp    mRNA            EST       03-AUG-1998
DEFINITION  AU014463 Schizosaccharomyces pombe late log phase cDNA
            Schizosaccharomyces pombe cDNA clone spc09897, mRNA sequence.
ACCESSION   AU014463
VERSION     AU014463.1  GI:3369254
KEYWORDS    EST.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   Morimyo,M. and Mita,K.
  TITLE     Identification of expressed sequence tags of Schizosaccharomyces
            pombe
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mitsuoki Morimyo
            Genome Research Group
            National Institute of Radiological Sciences
            9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
            Email: morimyo@nirs.go.jp.
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="Schizosaccharomyces pombe"
                     /strain="972"
                     /db_xref="taxon:4896"
                     /clone="spc09897"
                     /clone_lib="Schizosaccharomyces pombe late log phase cDNA"
                     /sex="h minus"
                     /note="Vector: M13mp19; The cDNA library of
                     Schizosaccharomyces pombe was prepared by cloning cDNA
                     into the SmaI site of M13mp19 DNA and the direction of DNA
                     sequences was not always from 5' to 3'. The cDNA data of
                     Schizosaccharomyces pombe are available for searching on
                     the World Wide Web.  (URL, http://www.nirs.go.jp)"
BASE COUNT       21 a      1 c      5 g      8 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 10;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 7.2e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gaaaaaaatta 13
           ||||||||||||
Db      16 GAAAAAAATTA 26


RESULT  38
D19995/c
LOCUS       D19995         35 bp    mRNA            EST       30-JUL-1996
DEFINITION  HUMGS00963 Human promyelocyte Homo sapiens cDNA clone mp1322 3',
            mRNA sequence.
ACCESSION   D19995
VERSION     D19995.1  GI:500892
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y., Yoshinari
            ,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60 before and
            after induction of differentiation. A new application of 3'directed
            cDNA sequencing
  JOURNAL   Unpublished (1993)
COMMENT     Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1 Yamada-oka,Suita,Osaka 565,Japan.
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="mp1322"
                     /clone_lib="Human promyelocyte"
                     /note="Female, adult, cell_line = HL60, cell_type =
```

promyelocyte. "
BASE COUNT 14 a 7 c 2 g 12 t
ORIGIN

Query Match 12.2%; Score 11; DB 11; Length 35;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 tgtaattttta 37
||||||||||||
Db 20 TGTAATTTTTA 10

RESULT 39
AU111149/c
LOCUS AU111149 36 bp mRNA EST 19-OCT-2000
DEFINITION AU111149 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk724e1 5', mRNA sequence.
ACCESSION AU111149
VERSION AU111149.1 GI:10924716
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 36)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S.
TITLE A complementary view of the C. elegans genome
JOURNAL Unpublished (2000)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .36
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk724e1"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 17 a 3 c 6 g 10 t
ORIGIN

Query Match 12.2%; Score 11; DB 10; Length 36;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aaaaattattt 16
||||||||||||
Db 23 AAAAATTATTT 13

RESULT 40
AZ381596/c
LOCUS AZ381596 36 bp DNA GSS 02-OCT-2000
DEFINITION 1M0138C16F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0138C16 F, DNA sequence.
ACCESSION AZ381596
VERSION AZ381596.1 GI:10495296
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 36)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0138 row: C column: 16
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 36.
FEATURES Location/Qualifiers
source 1. .36
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0138C16"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 19 a 2 c 10 g 5 t
ORIGIN

Query Match 12.2%; Score 11; DB 13; Length 36;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 catgttttctt 54
||||||||||||
Db 36 CATGTTTTCTT 26

RESULT 41
AZ807406/c
LOCUS AZ807406 36 bp DNA GSS 20-FEB-2001
DEFINITION 2M0070D15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0070D15 F, DNA sequence.
ACCESSION AZ807406
VERSION AZ807406.1 GI:12971722
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0070  row: D  column: 15
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 36.
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0070D15"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       18 a      7 c      7 g      4 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 7.2e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaattttt 36
           ||||||||||| 
Db      19 CTGTAATTTTT 9


RESULT  42
AA647854
LOCUS       AA647854       37 bp    mRNA            EST       28-OCT-1997
DEFINITION  vq80e05.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            IMAGE:1108640 5' similar to TR:G1136390 G1136390 KIAA0164 PROTEIN.
            ;, mRNA sequence.
ACCESSION   AA647854
VERSION     AA647854.1  GI:2574283
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:606808
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1108640"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage  embryos.  Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.   cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe  vector using commercial linkers (NEB).
                     Average insert size:  1.2 kb."
BASE COUNT       9 a      8 c      8 g     12 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 10;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 7.1e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      66 aataagtccaa 76
           ||||||||||| 
Db      26 AATAAGTCCAA 36


RESULT  43
AI200438
LOCUS       AI200438       37 bp    mRNA            EST       29-NOV-1998
DEFINITION  qf93b01.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA
            clone IMAGE:1757545 3' similar to SW:QPCT_HUMAN Q16769
            GLUTAMINYL-PEPTIDE CYCLOTRANSFERASE PRECURSOR ;contains element
            MER35 repetitive element ;, mRNA sequence.
ACCESSION   AI200438
VERSION     AI200438.1  GI:3753044
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 863   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1757545"
                     /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
                     /dev_stage="two placentae: one from 8 weeks and another
                     from 9 weeks post conception"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
BASE COUNT       11 a      6 c     12 g      8 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 10;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 7.1e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 tgaaaaaaatt 12
           ||||||||||
Db      14 TGAAAAAAATT 24


RESULT  44
AZ806836
LOCUS       AZ806836       37 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0069I08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0069I08 F, DNA sequence.
ACCESSION   AZ806836
VERSION     AZ806836.1  GI:12970584
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0069  row: I  column: 08
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 37.
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0069I08"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       14 a      1 c     10 g     12 t
ORIGIN


  Query Match             12.2%;  Score 11;  DB 13;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 7.1e+05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      30 aatttttatgt 40
           ||||||||||
Db       5 AATTTTTATGT 15


RESULT  45
D21038/c
LOCUS       D21038         38 bp    mRNA            EST       30-JUL-1996
DEFINITION  HUMGS02021 Human promyelocyte Homo sapiens cDNA clone mp0144 3',
            mRNA sequence.
ACCESSION   D21038
VERSION     D21038.1  GI:504858
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y., Yoshinari
            ,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60 before and
            after induction of differentiation. A new application of 3'directed
            cDNA sequencing
  JOURNAL   Unpublished (1993)
COMMENT     Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1 Yamada-oka,Suita,Osaka 565,Japan.
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="mp0144"
                     /clone_lib="Human promyelocyte"
                     /note="Female, adult, cell_line = HL60, cell_type =
                     promyelocyte. "
BASE COUNT       15 a      4 c      8 g     11 t
```

ORIGIN

Query Match 12.2%; Score 11; DB 11; Length 38;
Best Local Similarity 100.0%; Pred. No. 7e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      11 ttatttcaaag 21
           ||||||||||| 
Db      14 TTATTTCAAAG 4
```

Search completed: January 24, 2002, 02:57:32
Job time: 3954 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 03:22:36 ; Search time 1494.92 Seconds
(without alignments)
993.194 Million cell updates/sec

Title: US-09-531-438-4
Perfect score: 90
Sequence: 1 atgaaaaaaattatttcaaa.....gtccaatgaaagcaagtgca 90

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 541028

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	13	14.4	24	6	AX164353	AX164353 Sequence
2	13	14.4	26	6	AX039624	AX039624 Sequence
3	13	14.4	26	6	AX039654	AX039654 Sequence
4	13	14.4	30	6	AR028230	AR028230 Sequence
5	13	14.4	30	6	AR138633	AR138633 Sequence
6	13	14.4	33	5	XELARSE59	K01606 Xenopus lae
7	13	14.4	36	6	AX167671	AX167671 Sequence
c 8	13	14.4	39	6	AR011299	AR011299 Sequence
c 9	13	14.4	39	6	I17937	I17937 Sequence 16
10	12	13.3	14	6	AR082357	AR082357 Sequence
11	12	13.3	14	6	AR120899	AR120899 Sequence
12	12	13.3	14	6	I78403	I78403 Sequence 20
c 13	12	13.3	16	6	A14957	A14957 synthetic o
14	12	13.3	18	6	A14956	A14956 synthetic o
15	12	13.3	18	6	AR099365	AR099365 Sequence
16	12	13.3	20	6	A07597	A07597 Synthetic a
c 17	12	13.3	20	6	A64909	A64909 Sequence 66
c 18	12	13.3	21	6	AR043539	AR043539 Sequence
c 19	12	13.3	21	6	AR063857	AR063857 Sequence
c 20	12	13.3	21	6	AR075821	AR075821 Sequence
c 21	12	13.3	21	6	AR098733	AR098733 Sequence
c 22	12	13.3	21	6	AR112325	AR112325 Sequence
c 23	12	13.3	21	6	E30454	E30454 Method for
c 24	12	13.3	21	6	E32364	E32364 Method for
c 25	12	13.3	21	6	E33635	E33635 Detection o
c 26	12	13.3	21	6	E35692	E35692 Detection a
c 27	12	13.3	23	6	AR037890	AR037890 Sequence
28	12	13.3	24	6	AR054522	AR054522 Sequence
29	12	13.3	24	6	AR151501	AR151501 Sequence
c 30	12	13.3	25	6	AR082296	AR082296 Sequence
c 31	12	13.3	25	6	AR120838	AR120838 Sequence
32	12	13.3	25	6	AX115268	AX115268 Sequence
c 33	12	13.3	25	6	I78342	I78342 Sequence 14
c 34	12	13.3	27	6	AR060384	AR060384 Sequence
c 35	12	13.3	27	6	AR117878	AR117878 Sequence
c 36	12	13.3	27	6	AX114027	AX114027 Sequence
37	12	13.3	28	6	AR082955	AR082955 Sequence
c 38	12	13.3	28	6	AR082956	AR082956 Sequence
c 39	12	13.3	29	6	AX155921	AX155921 Sequence
c 40	12	13.3	30	6	A43687	A43687 Sequence 3
c 41	12	13.3	30	6	AR028176	AR028176 Sequence
c 42	12	13.3	30	6	AR050260	AR050260 Sequence
c 43	12	13.3	30	6	AR138579	AR138579 Sequence
c 44	12	13.3	30	6	AR140326	AR140326 Sequence
c 45	12	13.3	30	6	AX020976	AX020976 Sequence

ALIGNMENTS

RESULT 1
AX164353
LOCUS AX164353 24 bp DNA PAT 22-JUN-2001
DEFINITION Sequence 183 from Patent WO0138564.
ACCESSION AX164353
VERSION AX164353.1 GI:14545287
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 24)
AUTHORS Rouleau,G.A., Lafreniere,R.G., Rochefort,D., Cossette,P. and Ragsdale,D.
TITLE Loci for idiopathic generalized epilepsy, mutations thereof and method using same to assess, diagnose, prognose or treat epilepsy
JOURNAL Patent: WO 0138564-A 183 31-MAY-2001;
McGill University (CA)
FEATURES Location/Qualifiers
source 1. .24
/organism="synthetic construct"

```
                     /db_xref="taxon:32630"
                     /note="synthetic oligonucleotide"
BASE COUNT         2 a      5 c      3 g     14 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      34 tttatgttttcat 46
           |||||||||||||
Db       8 TTTATGTTTTCAT 20


RESULT   2
AX039624
LOCUS       AX039624       26 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 13 from Patent WO0063441.
ACCESSION   AX039624
VERSION     AX039624.1  GI:11229653
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Herrnstadt,C. and Davis,R.E.
  TITLE     Single nucleotide polymorphisms in mitochondrial genes that segreg
            ate with alzheimer's disease
  JOURNAL   Patent: WO 0063441-A 13 26-OCT-2000;
            MITOKOR (US)
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
BASE COUNT         8 a      2 c      5 g     11 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 atgttttcatgtt 49
           |||||||||||||
Db       7 ATGTTTTCATGTT 19


RESULT   3
AX039654
LOCUS       AX039654       26 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 43 from Patent WO0063441.
ACCESSION   AX039654
VERSION     AX039654.1  GI:11229683
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Herrnstadt,C. and Davis,R.E.
  TITLE     Single nucleotide polymorphisms in mitochondrial genes that segreg
            ate with alzheimer's disease
  JOURNAL   Patent: WO 0063441-A 43 26-OCT-2000;
            MITOKOR (US)
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
BASE COUNT         8 a      2 c      5 g     11 t
ORIGIN
```

```
  Query Match             14.4%;  Score 13;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 atgttttcatgtt 49
           |||||||||||||
Db       7 ATGTTTTCATGTT 19


RESULT   4
AR028230
LOCUS       AR028230       30 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 79 from patent US 5858661.
ACCESSION   AR028230
VERSION     AR028230.1  GI:5940203
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Ataxia-telangiectasia gene and its genomic organization
  JOURNAL   Patent: US 5858661-A 79 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT        13 a      2 c      1 g     14 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttc 17
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT   5
AR138633
LOCUS       AR138633       30 bp    DNA             PAT       16-JUN-2001
DEFINITION  Sequence 158 from patent US 6200749.
ACCESSION   AR138633
VERSION     AR138633.1  GI:14480978
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Mutated forms of the ataxia-telangiectasia gene and method to
            screen for a partial A-T phenotype
  JOURNAL   Patent: US 6200749-A 158 13-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT        13 a      2 c      1 g     14 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttc 17
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT   6
XELARSE59
```

```
LOCUS       XELARSE59      33 bp    DNA             VRT       28-APR-1993
DEFINITION  Xenopus laevis autonomous replication sequence e59.
ACCESSION   K01606
VERSION     K01606.1  GI:213953
KEYWORDS    autonomous replication; mutational analysis.
SOURCE      Xenopus laevis DNA.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Kearsey,S.
  TITLE     Structural requirements for the function of a yeast chromosomal
            replicator
  JOURNAL   Cell 37, 299-307 (1984)
  MEDLINE   84205653
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
BASE COUNT        8 a      4 c      2 g     19 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 5;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 atttttatgtttt 43
           |||||||||||||
Db       1 ATTTTTATGTTTT 13


RESULT   7
AX167671
LOCUS       AX167671       36 bp    DNA             PAT       03-JUL-2001
DEFINITION  Sequence 16 from Patent WO0144277.
ACCESSION   AX167671
VERSION     AX167671.1  GI:14597058
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   wegrich Glover,L., Budziszewski,G.J., Levin,J.Z. and Zhou,Q.
  TITLE     Herbicide target genes and methods
  JOURNAL   Patent: WO 0144277-A 16 21-JUN-2001;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide"
BASE COUNT        6 a      4 c     11 g     15 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      36 tatgttttcatgt 48
           |||||||||||||
Db      21 TATGTTTTCATGT 33


RESULT   8
AR011299/c
LOCUS       AR011299       39 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 168 from patent US 5762938.
ACCESSION   AR011299
VERSION     AR011299.1  GI:3969289
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Paoletti,E., Perkus,M.E., Taylor,J., Tartaglia,J., Norton,E.K.,
            Riviere,M., de Taisne,C., Limbach,K.J., Johnson,G.P., Pincus,S.E.,
            Cox,W.I., Audonnet,J.Francis and Gettig,R.Robert.
  TITLE     Modified recombinant vaccinia virus and expression vectors thereof
  JOURNAL   Patent: US 5762938-A 168 09-JUN-1998;
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="unknown"
BASE COUNT       15 a      6 c      7 g     11 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 6;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT   9
I17937/c
LOCUS       I17937         39 bp    DNA             PAT       07-OCT-1996
DEFINITION  Sequence 168 from patent US 5494807.
ACCESSION   I17937
VERSION     I17937.1  GI:1598292
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Paoletti,E., Perkus,M.E., Taylor,J., Tartaglia,J., Norton,E.K.,
            Riviere,M., de Taisne,C., Limbach,K.J., Johnson,G.P., Pincus,S.E.,
            Cox,W.I., Audonnet,J.F. and Gettig,R.R.
  TITLE     NYVAC vaccinia virus recombinants comprising heterologous inserts
  JOURNAL   Patent: US 5494807-A 168 27-FEB-1996;
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="unknown"
BASE COUNT       15 a      6 c      7 g     11 t
ORIGIN


  Query Match             14.4%;  Score 13;  DB 6;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaatttttat 38
           |||||||||||||
Db      25 CTGTAATTTTTAT 13


RESULT  10
AR082357
LOCUS       AR082357       14 bp    DNA             PAT       31-AUG-2000
DEFINITION  Sequence 201 from patent US 5972704.
ACCESSION   AR082357
VERSION     AR082357.1  GI:10009083
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and
            Thompson,J.D.
  TITLE     HIV nef targeted ribozymes
  JOURNAL   Patent: US 5972704-A 201 26-OCT-1999;
FEATURES             Location/Qualifiers
```

```
     source          1. .14
                     /organism="unknown"
BASE COUNT        7 a      3 c      3 g      1 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 CAATGAAAGCAA 13


RESULT  11
AR120899
LOCUS       AR120899       14 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 201 from patent US 6159692.
ACCESSION   AR120899
VERSION     AR120899.1  GI:14104475
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and
            Thompson,J.D.
  TITLE     Method and reagent for inhibiting human immunodeficiency virus
            replication
  JOURNAL   Patent: US 6159692-A 201 12-DEC-2000;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        7 a      3 c      3 g      1 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 CAATGAAAGCAA 13


RESULT  12
I78403
LOCUS       I78403         14 bp    DNA             PAT       03-APR-1998
DEFINITION  Sequence 201 from patent US 5693535.
ACCESSION   I78403
VERSION     I78403.1  GI:3014557
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and
            Thompson,J.D.
  TITLE     HIV targeted ribozymes
  JOURNAL   Patent: US 5693535-A 201 02-DEC-1997;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        7 a      3 c      3 g      1 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 CAATGAAAGCAA 13


RESULT  13
A14957/c
LOCUS       A14957         16 bp    DNA             PAT       18-FEB-1994
DEFINITION  synthetic oligonucleotide (N2) from patent EP0211299.
ACCESSION   A14957
VERSION     A14957.1  GI:491869
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Habermann,P., Stengelin,S. and Wengenmayer,F.
  TITLE     Fusion proteins, method for their production and their use
  JOURNAL   Patent: EP 0211299-A 2 25-FEB-1987;
            HOECHST AKTIENGESELLSCHAFT
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT        1 a      4 c      2 g      9 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db      15 CAATGAAAGCAA 4


RESULT  14
A14956
LOCUS       A14956         18 bp    DNA             PAT       18-FEB-1994
DEFINITION  synthetic oligonucleotide (N1) from patent EP0211299.
ACCESSION   A14956
VERSION     A14956.1  GI:491868
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Habermann,P., Stengelin,S. and Wengenmayer,F.
  TITLE     Fusion proteins, method for their production and their use
  JOURNAL   Patent: EP 0211299-A 1 25-FEB-1987;
            HOECHST AKTIENGESELLSCHAFT
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT        9 a      3 c      5 g      1 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       4 CAATGAAAGCAA 15


RESULT  15
AR099365
LOCUS       AR099365       18 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 19 from patent US 6077709.
```

```
ACCESSION   AR099365
VERSION     AR099365.1  GI:12809131
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Bennett,C.Frank, Ackermann,E.J., Swayze,E.E. and Cowsert,L.M.
  TITLE     Antisense modulation of Survivin expression
  JOURNAL   Patent: US 6077709-A 19 20-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unknown"
BASE COUNT        1 a      3 c      3 g     11 t
ORIGIN


  Query Match            13.3%;  Score 12;  DB 6;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      51 tcttattgttgg 62
           ||||||||||||
Db       1 TCTTATTGTTGG 12


RESULT  16
A07597
LOCUS       A07597         20 bp    DNA             PAT       09-JUL-1993
DEFINITION  Synthetic antisense oligonucleotide (5493-5512).
ACCESSION   A07597
VERSION     A07597.1  GI:413100
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Stropp,U., Baumgarten,J., Loebberding,A., Springer,W., Piel,N.,
            Kretschmer,A., Koelbl,H. and Frommer,W.
  TITLE     Antisense-oligonucleotides for inhibiting the transactivator target
            sequence (TAR) and the synthesis of the transactivator protein
            (Tat) of HIV-I, and their use
  JOURNAL   Patent: EP 0386563-A 4 12-SEP-1990;
            BAYER AG
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT        8 a      4 c      3 g      5 t
ORIGIN


  Query Match            13.3%;  Score 12;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db       2 CAATGAAAGCAA 13


RESULT  17
A64909/c
LOCUS       A64909         20 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 66 from Patent WO9731114.
ACCESSION   A64909
VERSION     A64909.1  GI:4530900
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Burnham,M.K. and Hodgson,J.E.
  TITLE     POLYNUCLEOTIDES AND AMINOACID SEQUENCES FROM STAPHYLOCOCCUS AUREUS
  JOURNAL   Patent: WO 9731114-A 66 28-AUG-1997;
            SMITHKLINE BEECHAM PLC (GB)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       11 a      4 c      2 g      3 t
ORIGIN


  Query Match            13.3%;  Score 12;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      20 GTTTTCATGTTT 9


RESULT  18
AR043539/c
LOCUS       AR043539       21 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 6 from patent US 5814490.
ACCESSION   AR043539
VERSION     AR043539.1  GI:5964547
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Spears,P.A.
  TITLE     Amplification and detection of chlamydia trachomatis nucleic acids
  JOURNAL   Patent: US 5814490-A 6 29-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match            13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  19
AR063857/c
LOCUS       AR063857       21 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 4 from patent US 5846726.
ACCESSION   AR063857
VERSION     AR063857.1  GI:5993165
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Nadeau,J.G., Pitner,J.Bruce, Schram,J.L., Linn,C.Preston, Vonk,G.P.
            and Walker,G.Terrance.
  TITLE     Detection of nucleic acids by fluorescence quenching
  JOURNAL   Patent: US 5846726-A 4 08-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match            13.3%;  Score 12;  DB 6;  Length 21;
```

```
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  20
AR075821/c
LOCUS       AR075821        21 bp    DNA             PAT       30-AUG-2000
DEFINITION  Sequence 4 from patent US 5958700.
ACCESSION   AR075821
VERSION     AR075821.1  GI:10002567
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Nadeau,J.G., Pitner,J.Bruce, Linn,C.Preston and Schram,J.L.
  TITLE     Detection of nucleic acids by fluorescence quenching
  JOURNAL   Patent: US 5958700-A 4 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  21
AR098733/c
LOCUS       AR098733        21 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 8 from patent US 6077669.
ACCESSION   AR098733
VERSION     AR098733.1  GI:12808499
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Little,M.C. and Vonk,G.P.
  TITLE     Kit and method for fluorescence based detection assay
  JOURNAL   Patent: US 6077669-A 8 20-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  22
AR112325/c
LOCUS       AR112325        21 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 5 from patent US 6130047.
ACCESSION   AR112325
VERSION     AR112325.1  GI:14092225
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Nadeau,J.G., Hsieh,H.V., Pitner,J.Bruce and Linn,C.Preston.
  TITLE     Detection of nucleic acids by fluorescence quenching
  JOURNAL   Patent: US 6130047-A 5 10-OCT-2000;
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  23
E30454/c
LOCUS       E30454          21 bp    DNA             PAT       07-FEB-2001
DEFINITION  Method for detecting target nucleic acid sequence and
            oligonucleotide.
ACCESSION   E30454
VERSION     E30454.1  GI:13025611
KEYWORDS    JP 1999056380-A/4.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   James,G.N.J.J., Pitona,C.P.R.R. and L,S.
  TITLE     Method for detecting target nucleic acid sequence and
  JOURNAL   Patent: JP 1999056380-A 4 02-MAR-1999;
            BECTON DICKINSON & CO
COMMENT     OS   Unidentified
            PN   JP 1999056380-A/4
            PD   02-MAR-1999
            PF   29-MAY-1998 JP 1998166141
            PR   30-MAY-1997 US   08/865.675
            PI   JAMES G NADEAU,J BLUCE PITONA,C PRESTON RIN,JAMES L SHURAMU PC
              C12N15/09,C12Q1/68,G01N33/50,G01N33/566,C12N15/00 CC
            Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .21
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  24
E32364/c
LOCUS       E32364          21 bp    DNA             PAT       07-FEB-2001
```

```
DEFINITION  Method for detecting nucleic acid by fluorescent quenching.
ACCESSION   E32364
VERSION     E32364.1  GI:13026696
KEYWORDS    JP 1999123083-A/4.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   James,G.N.J.J., Pitona,J.L.S.S., Preston,R.G.P.P. and G,T.W.
  TITLE     Method for detecting nucleic acid by fluorescent quenching
  JOURNAL   Patent: JP 1999123083-A 4 11-MAY-1999;
            BECTON DICKINSON & CO
COMMENT     OS   Unidentified
            PN   JP 1999123083-A/4
            PD   11-MAY-1999
            PF   11-MAY-1998 JP 1998127068
            PR   13-MAY-1997 US   08/855 085
            PI   JAMES G NADEAU,J BLUCE PITONA,JAMES L SHURAMU,C PRESTON RIN,
            PI   GREN P VONG,
            PI   G TERANSU WALKER
            PC   C12N15/09,C07H21/00,C12Q1/68,G01N21/64,G01N33/58,C12N15/00 CC
             Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .21
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  25
E33635/c
LOCUS       E33635          21 bp    DNA             PAT       07-FEB-2001
DEFINITION  Detection of nucleic acid by disappearance of fluorescence.
ACCESSION   E33635
VERSION     E33635.1  GI:13027030
KEYWORDS    JP 1999155598-A/5.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   James,G.N.H.H., C,J.B.P.P. and Preston,R.
  TITLE     Detection of nucleic acid by disappearance of fluorescence
  JOURNAL   Patent: JP 1999155598-A 5 15-JUN-1999;
            BECTON DICKINSON & CO
COMMENT     OS   Artificial Sequence
            PN   JP 1999155598-A/5
            PD   15-JUN-1999
            PF   22-SEP-1998 JP 1998267492
            PR   23-SEP-1997 US    08/933749
            PI   JAMES G NADEAU,HELEN V C,J BLUCE PITONA,C PRESTON RIN PC
            C12Q1/68,C07H21/00,C12N15/09,G01N21/64//G01N33/542,G01N33/566, PC
            C12N15/00
            CC
            FH   Key             Location/Qualifiers
            FT   source          1. .21
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      5 c      3 g      5 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  26
E35692/c
LOCUS       E35692          21 bp    DNA             PAT       07-FEB-2001
DEFINITION  Detection assay with the use of fluorescence and kit therefor.
ACCESSION   E35692
VERSION     E35692.1  GI:13019164
KEYWORDS    JP 1999225799-A/8.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Michael,C.L.G.G. and Vong.
  TITLE     Detection assay with the use of fluorescence and kit therefor
  JOURNAL   Patent: JP 1999225799-A 8 24-AUG-1999;
            BECTON DICKINSON & CO
COMMENT     OS   Artificial Sequence
            PN   JP 1999225799-A/8
            PD   24-AUG-1999
            PF   04-NOV-1998 JP 1998312790
            PR   04-NOV-1997 US    08/964020
            PI   MICHAEL C LITTLE,GREN P VONG
            PC   C12Q1/68,G01N21/78,G01N33/50//C12N15/09,C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   source          1. .21
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      5 c      3 g      5 t.
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 gttttcatgttt 50
           ||||||||||||
Db      13 GTTTTCATGTTT 2


RESULT  27
AR037890/c
LOCUS       AR037890        23 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 10 from patent US 5804383.
ACCESSION   AR037890
VERSION     AR037890.1  GI:5956607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Gruenert,D.C. and Dohrman,A.F.
  TITLE     Method and assay for detection of the expression of allele-specific
            mutations by allele-specific in situ reverse transcriptase
            polymerase chain reaction
  JOURNAL   Patent: US 5804383-A 10 08-SEP-1998;
FEATURES             Location/Qualifiers
```

```
     source          1. .23
                     /organism="unknown"
BASE COUNT        4 a      4 c      3 g     12 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      59 ttggagcaataa 70
           ||||||||||||
Db      23 TTGGAGCAATAA 12


RESULT  28
AR054522
LOCUS       AR054522       24 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 31 from patent US 5837441.
ACCESSION   AR054522
VERSION     AR054522.1  GI:5980099
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Hjelle,B. and Jenison,S.
  TITLE     Hantavirus-associated respiratory distress virus antigens
  JOURNAL   Patent: US 5837441-A 31 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        8 a      1 c      7 g      8 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 gtttactgtaat 32
           ||||||||||||
Db       2 GTTTACTGTAAT 13


RESULT  29
AR151501
LOCUS       AR151501       24 bp    DNA             PAT       08-AUG-2001
DEFINITION  Sequence 26 from patent US 6232094.
ACCESSION   AR151501
VERSION     AR151501.1  GI:15117551
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Hansson,L., Stromqvist,M., Bergstrom,S., Hernell,O. and Tornell,J.
  TITLE     DNA encoding human .kappa. casein and process for obtaining the
            protein
  JOURNAL   Patent: US 6232094-A 26 15-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        5 a      3 c      2 g     14 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      35 ttatgttttcat 46
           ||||||||||||
Db      12 TTATGTTTTCAT 23


RESULT  30
AR082296/c
LOCUS       AR082296       25 bp    DNA             PAT       31-AUG-2000
DEFINITION  Sequence 140 from patent US 5972704.
ACCESSION   AR082296
VERSION     AR082296.1  GI:10009022
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and
            Thompson,J.D.
  TITLE     HIV nef targeted ribozymes
  JOURNAL   Patent: US 5972704-A 140 26-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="unknown"
BASE COUNT        6 a      4 c      5 g     10 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db      23 CAATGAAAGCAA 12


RESULT  31
AR120838/c
LOCUS       AR120838       25 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 140 from patent US 6159692.
ACCESSION   AR120838
VERSION     AR120838.1  GI:14104414
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and
            Thompson,J.D.
  TITLE     Method and reagent for inhibiting human immunodeficiency virus
            replication
  JOURNAL   Patent: US 6159692-A 140 12-DEC-2000;
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="unknown"
BASE COUNT        6 a      4 c      5 g     10 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db      23 CAATGAAAGCAA 12


RESULT  32
AX115268
LOCUS       AX115268       25 bp    DNA             PAT       11-MAY-2001
DEFINITION  Sequence 391 from Patent WO0129262.
ACCESSION   AX115268
VERSION     AX115268.1  GI:14032210
KEYWORDS    .
```

```
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Picoult-Newburg,L. and Pohl,M.
  TITLE     Genotyping reagents, kits and methods of use thereof
  JOURNAL   Patent: WO 0129262-A 391 26-APR-2001;
            Orchid BioSciences, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Primer"
BASE COUNT        8 a      4 c      3 g     10 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 aaaattatttca 18
          ||||||||||||
Db      3 AAAATTATTTCA 14


RESULT  33
I78342/c
LOCUS       I78342         25 bp    DNA             PAT       03-APR-1998
DEFINITION  Sequence 140 from patent US 5693535.
ACCESSION   I78342
VERSION     I78342.1  GI:3014496
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Draper,K.G., Chowrira,B., McSwiggen,J., Stinchcomb,D.T. and
            Thompson,J.D.
  TITLE     HIV targeted ribozymes
  JOURNAL   Patent: US 5693535-A 140 02-DEC-1997;
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="unknown"
BASE COUNT        6 a      4 c      5 g     10 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
           ||||||||||||
Db      23 CAATGAAAGCAA 12


RESULT  34
AR060384/c
LOCUS       AR060384       27 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 6 from patent US 5840568.
ACCESSION   AR060384
VERSION     AR060384.1  GI:5986834
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Pfreundschuh,M.
  TITLE     Hodgkin's disease associated molecules and uses thereof
  JOURNAL   Patent: US 5840568-A 6 24-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
BASE COUNT        5 a      7 c      4 g     11 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      77 tgaaagcaagtg 88
           ||||||||||||
Db      26 TGAAAGCAAGTG 15


RESULT  35
AR117878/c
LOCUS       AR117878       27 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 6 from patent US 6140464.
ACCESSION   AR117878
VERSION     AR117878.1  GI:14098784
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Pfreundschuh,M. and Rammensee,H.
  TITLE     Nonapeptides that bind a HLA-A2.1 molecule
  JOURNAL   Patent: US 6140464-A 6 31-OCT-2000;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
BASE COUNT        5 a      7 c      4 g     11 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      77 tgaaagcaagtg 88
           ||||||||||||
Db      26 TGAAAGCAAGTG 15


RESULT  36
AX114027/c
LOCUS       AX114027       27 bp    DNA             PAT       08-MAY-2001
DEFINITION  Sequence 6 from Patent EP1108432.
ACCESSION   AX114027
VERSION     AX114027.1  GI:14018204
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Pfreundschuh,M.I.
  TITLE     Method for identifying or isolating a molecule and molecules
            indentified thereby
  JOURNAL   Patent: EP 1108432-A 6 20-JUN-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        5 a      7 c      4 g     11 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      77 tgaaagcaagtg 88
```

```
          ||||||||||||
Db      26 TGAAAGCAAGTG 15


RESULT  37
AR082955
LOCUS       AR082955        28 bp    DNA             PAT       01-SEP-2000
DEFINITION  Sequence 20 from patent US 5976795.
ACCESSION   AR082955
VERSION     AR082955.1  GI:10009745
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Voytas,D.F. and Zou,S.
  TITLE     Retrotransposon and methods
  JOURNAL   Patent: US 5976795-A 20 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="unknown"
BASE COUNT        7 a      0 c      4 g     15 t      2 others
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 atttttatgttt 42
           ||||||||||||
Db       7 ATTTTTATGTTT 18


RESULT  38
AR082956/c
LOCUS       AR082956        28 bp    DNA             PAT       01-SEP-2000
DEFINITION  Sequence 21 from patent US 5976795.
ACCESSION   AR082956
VERSION     AR082956.1  GI:10009746
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Voytas,D.F. and Zou,S.
  TITLE     Retrotransposon and methods
  JOURNAL   Patent: US 5976795-A 21 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="unknown"
BASE COUNT       15 a      4 c      0 g      7 t      2 others
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 atttttatgttt 42
           ||||||||||||
Db      22 ATTTTTATGTTT 11


RESULT  39
AX155921/c
LOCUS       AX155921        29 bp    DNA             PAT       22-JUN-2001
DEFINITION  Sequence 164 from Patent WO0140474.
ACCESSION   AX155921
VERSION     AX155921.1  GI:14537028
KEYWORDS    .
SOURCE      Chlamydia sp.
  ORGANISM  Chlamydia sp.
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Probst,P., Bhatia,A., Skeiky,Y.A., Fling,S.P. and Scholler,J.
  TITLE     Compounds and methods for treatment and diagnosis of chlamydial
            infection
  JOURNAL   Patent: WO 0140474-A 164 07-JUN-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="Chlamydia sp."
                     /db_xref="taxon:35827"
BASE COUNT       12 a      4 c      4 g      9 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 ttttcttattgt 59
           ||||||||||||
Db      15 TTTTCTTATTGT 4


RESULT  40
A43687/c
LOCUS       A43687          30 bp    DNA             PAT       06-MAR-1997
DEFINITION  Sequence 3 from Patent WO9508642.
ACCESSION   A43687
VERSION     A43687.1  GI:2298875
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Sammes,P.G. and Garman,A.J.
  TITLE     NUCLEIC ACID DETECTION WITH ENERGY TRANSFER
  JOURNAL   Patent: WO 9508642-A 3 30-MAR-1995;
            ZENECA LTD (GB)
COMMENT     Other publication AU 7662794 950410
            Other publication GB 2283095 950426.
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       12 a      4 c      4 g     10 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaattttta 37
           ||||||||||||
Db      29 CTGTAATTTTTA 18


RESULT  41
AR028176/c
LOCUS       AR028176        30 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 25 from patent US 5858661.
ACCESSION   AR028176
VERSION     AR028176.1  GI:5940149
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Ataxia-telangiectasia gene and its genomic organization
  JOURNAL   Patent: US 5858661-A 25 12-JAN-1999;
FEATURES             Location/Qualifiers
```

```
     source          1. .30
                     /organism="unknown"
BASE COUNT        8 a      2 c      2 g     18 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatt 15
          ||||||||||||
Db     19 AAAAAAATTATT 8


RESULT  42
AR050260/c
LOCUS       AR050260       30 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 3 from patent US 5827653.
ACCESSION   AR050260
VERSION     AR050260.1  GI:5972985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Sammes,P.George and Garman,A.John.
  TITLE     Nucleic acid detection with energy transfer
  JOURNAL   Patent: US 5827653-A 3 27-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT       12 a      4 c      4 g     10 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ctgtaattttta 37
          ||||||||||||
Db     29 CTGTAATTTTTA 18


RESULT  43
AR138579/c
LOCUS       AR138579       30 bp    DNA             PAT       16-JUN-2001
DEFINITION  Sequence 104 from patent US 6200749.
ACCESSION   AR138579
VERSION     AR138579.1  GI:14480924
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Mutated forms of the ataxia-telangiectasia gene and method to
            screen for a partial A-T phenotype
  JOURNAL   Patent: US 6200749-A 104 13-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT        8 a      2 c      2 g     18 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 aaaaaaattatt 15
          ||||||||||||
Db     19 AAAAAAATTATT 8


RESULT  44
AR140326/c
LOCUS       AR140326       30 bp    DNA             PAT       16-JUN-2001
DEFINITION  Sequence 14 from patent US 6207455.
ACCESSION   AR140326
VERSION     AR140326.1  GI:14482822
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Chang,L.-J.
  TITLE     Lentiviral vectors
  JOURNAL   Patent: US 6207455-A 14 27-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT        7 a      5 c      6 g     12 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 caatgaaagcaa 85
          ||||||||||||
Db     28 CAATGAAAGCAA 17


RESULT  45
AX020976/c
LOCUS       AX020976       30 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 22 from Patent EP0928832.
ACCESSION   AX020976
VERSION     AX020976.1  GI:10044639
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Kelly,S.J., Weston,S.L. and Robertson,N.H.
  TITLE     Cystic fibrosis test based on the detection of mutations in the
            cfre gene by arms
  JOURNAL   Patent: EP 0928832-A 22 14-JUL-1999;
            ZENECA LTD (GB)
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="PCR Amplification Primer"
BASE COUNT        8 a      6 c      5 g     11 t
ORIGIN


  Query Match             13.3%;  Score 12;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      59 ttggagcaataa 70
          ||||||||||||
Db     17 TTGGAGCAATAA 6


Search completed: January 24, 2002, 03:22:38
Job time: 3815 sec
```

```
                     /db_xref="taxon:32630"
                     /note="synthetic oligonucleotide"
BASE COUNT         2 a      5 c      3 g     14 t
ORIGIN


  Query Match           14.4%;  Score 13;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      34 tttatgttttcat 46
           |||||||||||||
Db       8 TTTATGTTTTCAT 20


RESULT   2
AX039624
LOCUS       AX039624       26 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 13 from Patent WO0063441.
ACCESSION   AX039624
VERSION     AX039624.1  GI:11229653
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Herrnstadt,C. and Davis,R.E.
  TITLE     Single nucleotide polymorphisms in mitochondrial genes that segreg
            ate with alzheimer's disease
  JOURNAL   Patent: WO 0063441-A 13 26-OCT-2000;
            MITOKOR (US)
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
BASE COUNT         8 a      2 c      5 g     11 t
ORIGIN


  Query Match           14.4%;  Score 13;  DB 6;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 atgttttcatgtt 49
           |||||||||||||
Db       7 ATGTTTTCATGTT 19


RESULT   3
AX039654
LOCUS       AX039654       26 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 43 from Patent WO0063441.
ACCESSION   AX039654
VERSION     AX039654.1  GI:11229683
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Herrnstadt,C. and Davis,R.E.
  TITLE     Single nucleotide polymorphisms in mitochondrial genes that segreg
            ate with alzheimer's disease
  JOURNAL   Patent: WO 0063441-A 43 26-OCT-2000;
            MITOKOR (US)
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
BASE COUNT         8 a      2 c      5 g     11 t
ORIGIN


  Query Match           14.4%;  Score 13;  DB 6;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 atgttttcatgtt 49
           |||||||||||||
Db       7 ATGTTTTCATGTT 19


RESULT   4
AR028230
LOCUS       AR028230       30 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 79 from patent US 5858661.
ACCESSION   AR028230
VERSION     AR028230.1  GI:5940203
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Ataxia-telangiectasia gene and its genomic organization
  JOURNAL   Patent: US 5858661-A 79 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT        13 a      2 c      1 g     14 t
ORIGIN


  Query Match           14.4%;  Score 13;  DB 6;  Length 30;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Ga

Qy       5 aaaaaattatttc 17
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT   5
AR138633
LOCUS       AR138633       30 bp    DNA             PAT       16-JUN-2001
DEFINITION  Sequence 158 from patent US 6200749.
ACCESSION   AR138633
VERSION     AR138633.1  GI:14480978
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Shiloh,Y.
  TITLE     Mutated forms of the ataxia-telangiectasia gene and method to
            screen for a partial A-T phenotype
  JOURNAL   Patent: US 6200749-A 158 13-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="unknown"
BASE COUNT        13 a      2 c      1 g     14 t
ORIGIN


  Query Match           14.4%;  Score 13;  DB 6;  Length 30;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 aaaaaattatttc 17
           |||||||||||||
Db      15 AAAAAATTATTTC 27


RESULT   6
XELARSE59
```

```
  Query Match             4.9%;  Score 16;  DB 20;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 aaaataattattttta 204
           ||||||||||||||||
Db       4 aaaataattattttta 19
```

```
RESULT   4
AAT97603
ID   AAT97603 standard; DNA; 36 BP.
XX
AC   AAT97603;
XX
DT   30-APR-1998  (first entry)
XX
DE   Shigella dysenteriae delta-stxA allele PCR primer 13.
XX
KW   Delta-virG allele; delta-guaB-A allele; PCR; amplification; primer;
KW   delta-stxA allele; shigellosis; vaccine; ss.
XX
OS   Synthetic.
OS   Shigella dysenteriae.
XX
PN   WO9737685-A1.
XX
PD   16-OCT-1997.
XX
PF   09-APR-1997;   97WO-US05954.
XX
PR   09-APR-1996;   96US-0629600.
XX
PA   (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI   Levine MM,  Noriega FR;
XX
DR   WPI; 1997-512417/47.
XX
PT   Shigella mutants with mutation in guaB-A - used in vaccines against
PT   Shigellosis
XX
PS   Example 6; Page 57; 94pp; English.
XX
CC   This is a PCR primer used in the amplification of the Shigella
CC   dysenteriae 1 delta-stxA allele.  The delta-stxA allele was integrated
CC   into delta-guaB-A of delta-guaB-A, delta-virG S. dysenteriae 1, which
CC   inactivated the shiga toxin of this strain.  The mutant can be used in
CC   the preparation of vaccines such as, a live vector vaccine comprising
CC   a Shigella mutant, (which encodes and expresses a foreign antigen, and
CC   a pharmaceutically acceptable carrier) or a DNA mediated vaccine
CC   comprising the Shigella mutant (which also contains a plasmid which
CC   encodes and expresses a foreign antigen in a eukaryotic cell).  The
CC   vaccines can be used against Shigellosis.
XX
SQ   Sequence 36 BP; 11 A; 3 C; 10 G; 12 T; 0 other;
```

```
  Query Match             4.9%;  Score 16;  DB 18;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 aaaataattattttta 204
           ||||||||||||||||
Db      14 aaaataattattttta 29
```

```
RESULT   5
AAV90775
ID   AAV90775 standard; DNA; 39 BP.
XX
AC   AAV90775;
XX
DT   18-FEB-1999  (first entry)
XX
DE   Primer Y104F.
XX
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; primer; ss.
XX
OS   Synthetic.
XX
PN   WO9849314-A2.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-US08487.
XX
PR   14-OCT-1997;   97US-0061958.
PR   25-APR-1997;   97US-0045107.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim Disclosure; Page 194; 402pp; English.
XX
CC   The specification, which describes Helicobacter pylori antigenic
CC   proteins that are characterised by immunoreactivity with
CC   H. pylori-positive antisera. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori
CC   antigens are used to detect H. pylori-specific antibodies, for
CC   diagnosing infection or to confirm eradication of infection, and
CC   in vaccines to protect against H. pylori infection and related
CC   diseases (gastritis, peptic ulcer, gastric adenocarcinoma/lymphoma).
CC   The present primer is used in the course of the invention.
XX
SQ   Sequence 39 BP; 15 A; 8 C; 7 G; 9 T; 0 other;
```

```
  Query Match             4.9%;  Score 16;  DB 20;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 atatcttaaatttagc 23
           ||||||||||||||||
Db      23 atatcttaaatttagc 38
```

```
RESULT   6
AAZ68379
ID   AAZ68379 standard; DNA; 47 BP.
XX
AC   AAZ68379;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:2726.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,A)
```

SEQ ID alignments

```
  Query Match             4.9%;  Score 16;  DB 20;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 aaaataattattttta 204
           ||||||||||||||||
Db       4 aaaataattattttta 19
```

```
RESULT   4
AAT97603
ID   AAT97603 standard; DNA; 36 BP.
XX
AC   AAT97603;
XX
DT   30-APR-1998  (first entry)
XX
DE   Shigella dysenteriae delta-stxA allele PCR primer 13.
XX
KW   Delta-virG allele; delta-guaB-A allele; PCR; amplification; primer;
KW   delta-stxA allele; shigellosis; vaccine; ss.
XX
OS   Synthetic.
OS   Shigella dysenteriae.
XX
PN   WO9737685-A1.
XX
PD   16-OCT-1997.
XX
PF   09-APR-1997;   97WO-US05954.
XX
PR   09-APR-1996;   96US-0629600.
XX
PA   (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI   Levine MM,  Noriega FR;
XX
DR   WPI; 1997-512417/47.
XX
PT   Shigella mutants with mutation in guaB-A - used in vaccines against
PT   Shigellosis
XX
PS   Example 6; Page 57; 94pp; English.
XX
CC   This is a PCR primer used in the amplification of the Shigella
CC   dysenteriae 1 delta-stxA allele.  The delta-stxA allele was integrated
CC   into delta-guaB-A of delta-guaB-A, delta-virG S. dysenteriae 1, which
CC   inactivated the shiga toxin of this strain.  The mutant can be used in
CC   the preparation of vaccines such as, a live vector vaccine comprising
CC   a Shigella mutant, (which encodes and expresses a foreign antigen, and
CC   a pharmaceutically acceptable carrier) or a DNA mediated vaccine
CC   comprising the Shigella mutant (which also contains a plasmid which
CC   encodes and expresses a foreign antigen in a eukaryotic cell).  The
CC   vaccines can be used against Shigellosis.
XX
SQ   Sequence 36 BP; 11 A; 3 C; 10 G; 12 T; 0 other;
```

```
  Query Match             4.9%;  Score 16;  DB 18;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 aaaataattattttta 204
           ||||||||||||||||
Db      14 aaaataattattttta 29
```

```
RESULT   5
AAV90775
ID   AAV90775 standard; DNA; 39 BP.
XX
AC   AAV90775;
XX
DT   18-FEB-1999  (first entry)
XX
DE   Primer Y104F.
XX
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; primer; ss.
XX
OS   Synthetic.
XX
PN   WO9849314-A2.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-US08487.
XX
PR   14-OCT-1997;   97US-0061958.
PR   25-APR-1997;   97US-0045107.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim Disclosure; Page 194; 402pp; English.
XX
CC   The specification, which describes Helicobacter pylori antigenic
CC   proteins that are characterised by immunoreactivity with
CC   H. pylori-positive antisera. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori
CC   antigens are used to detect H. pylori-specific antibodies, for
CC   diagnosing infection or to confirm eradication of infection, and
CC   in vaccines to protect against H. pylori infection and related
CC   diseases (gastritis, peptic ulcer, gastric adenocarcinoma/lymphoma).
CC   The present primer is used in the course of the invention.
XX
SQ   Sequence 39 BP; 15 A; 8 C; 7 G; 9 T; 0 other;
```

```
  Query Match             4.9%;  Score 16;  DB 20;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 atatcttaaatttagc 23
           ||||||||||||||||
Db      23 atatcttaaatttagc 38
```

```
RESULT   6
AAZ68379
ID   AAZ68379 standard; DNA; 47 BP.
XX
AC   AAZ68379;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human map-related biallelic marker SEQ ID NO:2726.
XX
KW   Human genome; biallelic marker; high density disequilibrium map;
KW   genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW   haplotyping; hybridisation; identification; characterisation;
KW   diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(24,A)
```

100% identity

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 24, 2002, 02:27:08 ; Search time 222.28 Seconds
(without alignments)
1261.226 Million cell updates/sec

Title: US-09-531-438-3
Perfect score: 327
Sequence: 1 atttgggatatcttaaattt.....tttcatgttttcttattgtt 327

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621.seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 989696

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16	4.9	29	17	AAT42655	Primer for amplify
2	16	4.9	29	21	AAA51200	N-terminal primer
3	16	4.9	32	20	AAZ27689	PCR primer for Ver
4	16	4.9	36	18	AAT97603	Shigella dysenteri
5	16	4.9	39	20	AAV90775	Primer Y104F. Syn
6	16	4.9	47	21	AAZ68379	Human map-related
7	15	4.6	20	21	AAA66183	Dog genomic marker
8	15	4.6	24	22	AAH46804	Human high motilit
9	15	4.6	27	17	AAT39439	Hel-N2 selected se
10	15	4.6	27	19	AAV37457	Human Hel-N2 selec
11	15	4.6	30	14	AAQ43975	Triple helix formi
12	15	4.6	48	16	AAT25576	Human gene signatu
13	14	4.3	18	13	AAQ20161	Cross-linking olig
14	14	4.3	18	13	AAQ20160	Cross-linking olig
15	14	4.3	18	13	AAQ30311	Oligomer HSV724 fo
16	14	4.3	18	13	AAQ30310	Oligomer HSV723 fo
c 17	14	4.3	18	20	AAZ22162	Human c-IAP-1 mRNA
18	14	4.3	22	22	AAS01590	Human IQGAP2 CpG i
c 19	14	4.3	22	22	AAS01643	Human IQGAP2 5' ta
c 20	14	4.3	26	19	AAV07952	Helicobacter pylor
c 21	14	4.3	26	19	AAV07922	Helicobacter pylor
c 22	14	4.3	27	19	AAV07937	Helicobacter pylor
c 23	14	4.3	29	17	AAT42655	Primer for amplify
c 24	14	4.3	29	21	AAA51200	N-terminal primer
c 25	14	4.3	31	16	AAT25703	Human gene signatu
26	14	4.3	31	19	AAV67854	Nucleotide fragmen
c 27	14	4.3	32	20	AAZ27689	PCR primer for Ver
c 28	14	4.3	36	18	AAT97603	Shigella dysenteri
29	14	4.3	36	22	AAC90606	Tomato spotted wil
30	14	4.3	37	15	AAQ62952	Glycophorin antibo
c 31	14	4.3	45	22	AAF55449	Oligonucleotide us
32	14	4.3	45	22	AAF55450	Oligonucleotide us
c 33	14	4.3	45	22	AAC88874	Oligonucleotide TA
34	14	4.3	45	22	AAC88875	Oligonucleotide TA
35	14	4.3	47	21	AAZ66366	Human map-related
c 36	14	4.3	47	21	AAZ67473	Human map-related
c 37	14	4.3	47	21	AAZ67533	Human map-related
c 38	14	4.3	47	21	AAZ67549	Human map-related
39	14	4.3	47	21	AAZ67813	Human map-related
40	14	4.3	50	21	AAA98312	Human MSH6 fragmen
41	13	4.0	15	22	AAF48097	IGFBP3 oligonucleo
42	13	4.0	15	22	AAF48098	IGFBP3 oligonucleo
43	13	4.0	15	22	AAF48099	IGFBP3 oligonucleo
c 44	13	4.0	16	21	AAA57758	Nucleotide sequenc
c 45	13	4.0	17	16	AAQ92084	Renilla reniformis

ALIGNMENTS

RESULT 1
AAT42655
ID AAT42655 standard; DNA; 29 BP.
XX
AC AAT42655;
XX
DT 25-FEB-1997 (first entry)
XX
DE Primer for amplifying verotoxin (VT-1) subunit A coding sequence.
XX
KW Verotoxin; Escherichia coli; enteric infection; diarrhoea; vaccine;
KW haemolytic uraemic syndrome; detection; ss.
XX
OS Synthetic.
XX
PN WO9630043-A1.
XX
PD 03-OCT-1996.
XX
PF 25-MAR-1996; 96WO-US04093.
XX
PR 24-MAR-1995; 95US-0410058.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Carroll SB, Padhye NV, Stafford DC;
XX
DR WPI; 1996-505779/50.
XX
PT Compsn. contg. neutralising antitoxin against E.coli vero-toxin -
PT used to treat intoxicated individuals, and as a prophylactic against
PT diarrhoeal disease or extra-intestinal complications of E.coli
PT infection
XX

PS Example 6; Page 58; 101pp; English.
XX
CC Compositions containing neutralising antitoxin against one or more E.
CC coli verotoxin (VT) can be used to treat intoxicated adults and
CC children with enteric bacterial infections. They may also be used as
CC prophylactics e.g. as a vaccine, against diarrhoeal disease or the
CC development of extra-intestinal complications of E.coli infection,
CC especially haemolytic uraemic syndrome. The antitoxin can also be
CC used to detect E. coli VT in a sample. The VT is recombinant,
CC preferably a fusion protein containing a non-VT protein sequence and
CC part of the E.coli VT1 or VT2 sequence. Two primers (AAT42655,
CC AAT42656) were used to amplify the verotoxin VT-1 A subunit coding
CC sequence and add a histidine tag coding sequence to the subunit
CC sequence. Two primers (AAT42655, AAT42658) were used to amplify the
CC verotoxin VT-1 A and B subunits and add a histidine tag coding
CC sequence to the subunit sequences.
XX
SQ Sequence 29 BP; 11 A; 2 C; 5 G; 11 T; 0 other;

Query Match 4.9%; Score 16; DB 17; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 aaaataattattttta 204
||||||||||||||||
Db 9 aaaataattattttta 24

RESULT 2
AAA51200
ID AAA51200 standard; DNA; 29 BP.
XX
AC AAA51200;
XX
DT 26-SEP-2000 (first entry)
XX
DE N-terminal primer for E. coli verotoxin 1 subunit A gene.
XX
KW VT-1; verotoxin; antitoxin therapy; fusion protein; affinity tag; food;
KW recombinant production; screening; dairy; anti-bacterial; vaccine;
KW primer; polyhistidine; ss.
XX
OS Escherichia coli.
OS Synthetic.
XX
PN US6080400-A.
XX
PD 27-JUN-2000.
XX
PF 13-MAR-1997; 97US-0816977.
XX
PR 24-MAR-1995; 95US-0410058.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Byrne LM;
XX
DR WPI; 2000-451195/39.
XX
PT Bacterial cell for recombinantly expressing bacterial toxins in large
PT quantities useful for immunization and treatment of bacterial
PT infections, comprises expression vector encoding bacterial toxin
XX
PS Example 6; Column 83; 83pp; English.
XX
CC E. coli verotoxin (VT) type 1 and 2 subunits A and B were cloned into
CC pET-23b, designed to allow expression of the native proteins containing
CC C-terminal polyhistidine tags. The VT-1 and VT-2 genes were engineered
CC to convert the signal sequence methionine codon into a NdeI site to
CC allow cloning of the amplified genes into the vector without addition of
CC vector-encoded amino acids. The C-terminal primers comprises the
CC terminal 7 codons of each gene fused to the sequence CTCGAGCC, in order
CC to add the polyhistidine tag. The primers delete the native stop codons,
CC and when cloned into pET-23 add a C-terminal extension of Leu-Glu-(His)6.
CC VT B chains are small proteins (approximately 8 kDa), so use of a small
CC affinity tag was preferred (i.e. polyhistidine). A polyhistidine affinity
CC tag facilitates single step affinity purification of subunits from
CC periplasmic extracts. However, due to poor recovery of his-tagged VT-1 A
CC and VT-2 A chains, expression of maltose binding protein (MBP) fused
CC subunits was undertaken. Due to the toxicity of the VT-2 B subunit,
CC strict uninduced promoter control is necessary to permit cell viability.
CC Bacterial host cells expressing a recombinant expression vector encoding
CC a polyhistidine affinity tag and a portion of the VT-2 B chain are
CC claimed. The vector is chosen from pET24hisVT2BL+, pET24hisVT2BL- and
CC pET24VT2B, where "L+" indicates that the vector encodes the preprotein
CC form of the protein and "L-" indicates that the vector encodes the mature
CC form of the protein. The bacterial cell is capable of expressing large
CC quantities (40 mg/l) of VT-2B. The toxins are useful for immunizing
CC non-mammals and for detecting bacterial toxins in environmental samples
CC including soil, water, industrial samples, biological samples and samples
CC obtained from food and dairy processing instruments.
XX
SQ Sequence 29 BP; 11 A; 2 C; 5 G; 11 T; 0 other;

Query Match 4.9%; Score 16; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 aaaataattattttta 204
||||||||||||||||
Db 9 aaaataattattttta 24

RESULT 3
AAZ27689
ID AAZ27689 standard; DNA; 32 BP.
XX
AC AAZ27689;
XX
DT 22-DEC-1999 (first entry)
XX
DE PCR primer for Verotoxin gene.
XX
KW Verotoxin; VT1; VT2; detection; PCR primer; ss.
XX
OS Synthetic.
OS Escherichia coli.
XX
PN JP11243996-A.
XX
PD 14-SEP-1999.
XX
PF 27-FEB-1998; 98JP-0047677.
XX
PR 27-FEB-1998; 98JP-0047677.
XX
PA (TOYM) TOYOBO KK.
XX
DR WPI; 1999-603716/52.
XX
PT An oligonucleotide for amplification of verotoxin - useful in the
PT detection of inactivated verotoxin gene by transfer of a foreign DNA
PT fragment
XX
PS Claim 11; Page 9; 10pp; Japanese.
XX
CC This sequence represents a PCR primer of the invention. The primer is
CC used for amplification of the E. coli verotoxin (VT) gene. The
CC oligonucleotide is useful for detection of inactivated VT gene by
CC transfer of a foreign DNA fragment. Simple, rapid and specific
CC amplification of VT gene from environmental factors is achieved using the
CC oligonucleotide of the invention.
XX
SQ Sequence 32 BP; 12 A; 2 C; 4 G; 14 T; 0 other;

```
  Query Match             4.8%;  Score 15;  DB 16;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     159 aaaatataaataaaa 173
           ||||||||||||||||
Db      17 aaaatataaataaaa 31
```

```
RESULT  13
AAQ20161
ID   AAQ20161 standard; DNA; 18 BP.
XX
AC   AAQ20161;
XX
DT   01-APR-1992  (first entry)
XX
DE   Cross-linking oligomer 724 to target Herpes Simplex Virus I.
XX
KW   deoxyribonucleic acid; major groove; HSV;
KW   inverted polarity region; covalent cross-linking group; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag= a
FT                   /mod_base= OTHER
FT                   /note= "N4N4-ethanocytosine"
FT   modified_base   2
FT                   /*tag= b
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   3
FT                   /*tag= c
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   4
FT                   /*tag= d
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   5
FT                   /*tag= e
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   7
FT                   /*tag= f
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   9
FT                   /*tag= g
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   10
FT                   /*tag= h
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   11
FT                   /*tag= i
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   misc_feature    12..18
FT                   /*tag= j
FT                   /label= inverted_polarity_region
FT                   /note= "see comments"
FT   modified_base   13
FT                   /*tag= k
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   14
FT                   /*tag= l
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   15
FT                   /*tag= m
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   17
FT                   /*tag= n
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
XX
PN   WO9118997-A.
XX
PD   12-DEC-1991.
XX
PF   24-MAY-1991;   91WO-1003680.
XX
PR   14-JAN-1991;   91US-0640654.
PR   25-MAY-1990;   90US-0529346.
XX
PA   (GILE-) GILEAD SCIE INC.
XX
PI   Matteucci MD,  Krawczyk S;
XX
DR   WPI; 1992-007480/01.
XX
PT   New sequence-specific non-photo-activated crosslinking agents -
PT   bind to the major groove of duplex DNA and are esp. useful for
PT   treating latent infections e.g. HIV
XX
PS   Example 4; Page 29; 42pp; English.
XX
CC   This oligomer contains an inverted polarity region formed from an
CC   o-xyloso dimer synthon. Residues 11 and 12 are linked via an
CC   o-xyloso group (i.e. nucleotides that have xylose sugar linked via
CC   the o-xylene ring). The sequence is designed to target the Herpes
CC   Simplex virus I beginning at nucleotide 10996 and to covalently
CC   cross-link to it. See also AAQ20151-Q20160.
XX
SQ   Sequence 18 BP; 12 A; 1 C; 0 G; 5 T; 0 other;
```

```
  Query Match             4.3%;  Score 14;  DB 13;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     159 aaaatataaataaa 172
           ||||||||||||||
Db       2 aaaatataaataaa 15
```

```
RESULT  14
AAQ20160
ID   AAQ20160 standard; DNA; 18 BP.
XX
AC   AAQ20160;
XX
DT   01-APR-1992  (first entry)
XX
DE   Cross-linking oligomer 723 to target Herpes Simplex Virus I.
XX
KW   deoxyribonucleic acid; major groove; HSV;
KW   inverted polarity region; covalent cross-linking group; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag= a
FT                   /mod_base= OTHER
FT                   /note= "N-methyl-8-oxo-2'-deoxyadenine"
FT   modified_base   2
FT                   /*tag= b
FT                   /mod_base= OTHER
```